In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

University of Alberta

Structure and Dynamics of the *Tenebrio molitor* Antifreeze Protein

by

Margaret Elise Daley ©

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Department of Biochemistry

Edmonton, Alberta, Canada
Fall 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87958-5
Our file Notre référence
ISBN: 0-612-87958-5

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: Margaret Elise Daley

Title of Thesis: Structure and Dynamics of the *Tenebrio molitor* Antifreeze Protein

Degree: Doctor of Philosophy

Year this Degree Granted: 2003

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Date: September 22, 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled "Structure and Dynamics of the *Tenebrio molitor* Antifreeze Protein" submitted by Margaret Elise Daley in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Prof. Brian D. Sykes

Prof. Michael N. G. James

Prof. Andrew M. MacMillan

Prof. George Kotovych

Prof. Robert G. Bryant

September 16, 2003

To Mom, Dad, and Chris

ABSTRACT

Antifreeze proteins protect many types of organisms from damage caused by freezing. They do this by binding to the ice surface, which causes inhibition of ice crystal growth. However, the molecular mechanism of ice binding leading to growth inhibition is not well understood. In this thesis, NMR investigations of the structure and dynamics of the antifreeze protein from the yellow mealworm beetle *Tenebrio molitor* are presented. TmAFP is a small, highly disulfide-bonded, right-handed parallel β-helix consisting of seven tandemly repeated 12 amino acid loops. Analysis of the ^{15}N relaxation parameters shows that TmAFP is a well-defined, rigid structure, and the extracted parameters reveal similar restricted internal mobility throughout the protein backbone at both 30 and 5°C. In TmAFP, an array of threonine residues on one face of the protein is responsible for conferring its ability to bind crystalline ice and inhibit its growth. The flexibility of the threonine side chains was investigated using two-dimensional homonuclear NMR spectroscopy and natural abundance ^{13}C relaxation. From measurement of the $^{3}J_{\alpha\beta}$ ^{1}H-^{1}H scalar coupling constants, the χ_1 angles and preferred rotamer populations can be calculated. It was determined that the threonines on the ice-binding face of the protein adopt a preferred rotameric conformation at near freezing temperatures, whereas threonines not on the ice-binding face experience conformational averaging. This suggests that TmAFP maintains a preformed ice-binding conformation in solution, wherein the rigid array of threonines that form the AFP-ice interface matches the ice crystal lattice. The CαH relaxation measurements are compared to the measured ^{15}N backbone parameters and these are found to be in agreement. For analysis of the

threonine side chain motions, a model of restricted rotational diffusion about the χ_1 dihedral angle was employed. The range of motion experienced by the ice-binding threonine side chains is restricted, corresponding to a range of less than ±25°. A key factor in binding to the ice surface and inhibition of ice crystal growth appears to be the close surface-to-surface complementarity between the AFP and crystalline ice, and the lack of an entropic penalty associated with freezing out motions in a flexible ligand.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr. Brian Sykes, for his excellent guidance and support during my PhD studies. My time in this lab has been extremely rewarding and I have learned so much and enjoyed it greatly. I would like to thank all of my co-workers in the Sykes laboratory, past and present - the following people in particular: Steffen Graether and Carolyn Slupsky, my fellow members of 'Team AFP' for help and many valuable discussions; Gerry McQuaid for maintenance of the spectrometers; Robert Boyko, Leigh Willard, and Dean Schieve for their computer expertise; Stéphane Gagné for useful software programs and scripts; and Pascal Mercier, we were the only students when we were starting and survived the courses together and for his helpful NMRView scripts. I also thank Dr. Leo Spyracopoulos for his dynamics and pulse sequence expertise and for always being ready to answer my questions. Dr. Lewis Kay of the University of Toronto is acknowledged for pulse sequences.

I would also like to thank Dr. Peter Davies of Queen's University in Kingston, Ontario for our strong collaboration. Our emails and conversations over the years have been very valuable and receiving beautiful protein samples in the mail was wonderful. I appreciate the work of members of the Davies laboratory, in particular Yih-Cherng Liou and Sherry Gauthier for the preparation of the 'ultra-primo' TmAFP, as well as the ^{15}N-labeled sample and Chris Marshall for many helpful discussions and all of his work preparing the Tm 4-9 protein and the many mutations. I also thank Dr. Zongchao Jia of Queen's University for kindly providing the x-ray coordinates of TmAFP prior to their publication.

Funding for this work was provided by the Canadian Institutes of Health Research (CIHR) and the Protein Engineering Network of Centres of Excellence (PENCE). I would like to acknowledge the CIHR for my Doctoral Research Award and the University of Alberta for the Walter H. Johns Graduate Fellowship. I also thank the Canadian National High Field NMR Centre (NANUC) for their assistance and use of the facilities. The operation of NANUC is funded by the CIHR, the Natural Science and Engineering Research Council of Canada and the University of Alberta.

Finally I would like to thank my husband Chris and my family for all of the support and encouragement over the years, I would not have been able to do it without them!

TABLE OF CONTENTS

CHAPTER 1: General Introduction 1
- **1.1 Biological Significance** 1
- **1.2 Structural Diversity of AFPs** 3
- **1.3 Ice Binding Mechanism** 6
- **1.4 NMR Spectroscopy** 8
- **1.5 Objectives** 9
- **1.6 References** 11

CHAPTER 2: The Solution Structure and Backbone Dynamics of the *Tenebrio molitor* Antifreeze Protein 14
- **2.1 Introduction** 14
- **2.2 Experimental Procedures** 16
  - *2.2.1 Sample preparation* 16
  - *2.2.2 Assignment* 16
  - *2.2.3 Structure calculations* 18
  - *2.2.4 NMR spectroscopy for backbone dynamics* 18
  - *2.2.5 NMR data processing* 19
  - *2.2.6 ^{15}N-Relaxation data analysis* 20
- **2.3 Results** 24
  - *2.3.1 Solution structure of TmAFP* 24
  - *2.3.2 ^{15}N-T_1, -T_2, and NOE data* 28
  - *2.3.3 Rotational diffusion anisotropy and overall correlation time.* 31
  - *2.3.4 Model-free analysis and internal mobility* 34
- **2.4 Discussion** 35
- **2.5 References** 37

CHAPTER 3: The Role of Side Chain Conformational Flexibility in Surface Recognition by *Tenebrio molitor* Antifreeze Protein 41
- **3.1 Introduction** 41
- **3.2 Results** 43
- **3.3 Discussion** 52
- **3.4 Materials and Methods** 54
  - *3.4.1 NMR spectroscopy* 54
  - *3.4.2 Determination of $^3J_{\alpha\beta}$ coupling constants* 55
  - *3.4.3 Determination of χ_1 side chain torsion angles* 56
- **3.5 Serine and Valine Mutation Studies** 57
  - *3.5.1 Rationale* 57
  - *3.5.2 Experimental procedures* 58
  - *3.5.3 Results* 58
  - *3.5.4 Discussion* 60
- **3.6 References** 61

CHAPTER 4: Characterization of Threonine Side Chain Dynamics Using Natural Abundance ^{13}C NMR Spectroscopy....64
4.1 Introduction....64
4.2 Experimental Procedures....66
  4.2.1 NMR spectroscopy....66
  4.2.2 NMR data processing....68
4.3 Theory....69
4.4 Results....71
  4.4.1 ^{13}C-R_1, -R_2, and NOE data....71
  4.4.2 Estimation of the overall correlation time....77
4.5 Discussion....78
4.6 Conclusions....80
4.7 References....81

CHAPTER 5: The Temperature Dependence of Amide Proton Chemical Shifts....84
5.1 Introduction....84
5.2 Experimental Procedures....86
  5.2.1 NMR spectroscopy....86
5.3 Results and Discussion....87
5.4 Conclusions....93
5.5 References....93

CHAPTER 6: General Conclusions....95
References....99

APPENDICES....101
A Chemical shift assignments of TmAFP....101
B Assessment of TmAFP folding by NMR....105
C Identification of the TmAFP ice-binding face....110
D Assessment of engineered TmAFP with deleted and added loops....116
E Purification of TmAFP by inclusion into ice....121

LIST OF TABLES

Table 1.1: Antifreeze proteins from fish and insects.. 4
Table 2.1: Structural statistics..26
Table 2.2: Rotational diffusion parameters...33
Table 3.1: Percent population of χ_1 rotamer conformations..................................51
Table 3.2: Threonine coupling constants..59
Table 3.3: Threonine coupling constants for T40 mutations...................................59
Table 4.1: Threonine and valine ^{13}C R_1 measurements...76
Table 4.2: Threonine ^{13}C R_2 measurements...76
Table 4.3: Threonine and valine ^{13}C NOE measurements...77
Table 5.1: NH shift deviations and temperature gradients...................................... 89
Table A.1: TmAFP chemical shift assignments... 101

LIST OF FIGURES

Figure 1.1: Thermal hysteresis....3
Figure 1.2: AFP structural diversity.... 5
Figure 2.1: TmAFP amino acid sequence....17
Figure 2.2: ^{15}N-HSQC spectrum.... 24
Figure 2.3: Ensemble of structures.... 25
Figure 2.4: Ribbon representation of TmAFP structure.... 27
Figure 2.5: Comparison of crystal and solution structures.... 28
Figure 2.6: ^{15}N relaxation parameters at 30°C....29
Figure 2.7: ^{15}N relaxation parameters at 5°C....30
Figure 2.8: TmAFP rotational diffusion tensor....33
Figure 2.9: ^{15}N relaxation order parameters....34
Figure 3.1: Sequence and structure of TmAFP....44
Figure 3.2: ^{1}H NMR spectra in freezing experiment....45
Figure 3.3: 2D DQF-COSY spectrum.... 46
Figure 3.4: Expansions of 2D DQF-COSY spectra....47
Figure 3.5: Plot of $^{3}J_{\alpha\beta}$ vs temperature.... 49
Figure 3.6: Staggered rotamer conformations of threonine....50
Figure 3.7: Structure of Type I AFP....57
Figure 3.8: Plot of $^{3}J_{\alpha\beta}$ for mutations according to sequence....60
Figure 4.1: ^{13}C-HSQC spectrum.... 67
Figure 4.2: Restricted rotation around Cα-Cβ bond....69
Figure 4.3: ^{13}C relaxation parameters at 500 MHz and 30°C.... 72
Figure 4.4: ^{13}C relaxation parameters at 600 MHz and 30°C.... 73
Figure 4.5: ^{13}C relaxation parameters at 500 MHz and 5°C.... 74
Figure 4.6: ^{13}C relaxation parameters at 600 MHz and 5°C.... 75
Figure 4.7: Calculated graph for ^{13}C R_1.... 79
Figure 4.8: Calculated graph for ^{13}C R_2.... 79
Figure 4.9: Calculated graph for ^{13}C NOE....80
Figure 5.1: TmAFP ribbon structure....85
Figure 5.2: ^{15}N-HSQC spectra as a function of temperature....87
Figure 5.3: Amide chemical shifts mapped on TmAFP surface....88
Figure 5.4: Amide CSD-$\Delta\delta/\Delta T$ correlation plot....91
Figure B.1: ^{1}H NMR spectra of Tm 2-14....107
Figure B.2: Comparative TOCSY spectra of ‘fingerprint’ region....108
Figure C.1: Comparison of Tm 2-14 and 4-9 isoforms.... 112
Figure C.2: Comparison of ‘fingerprint’ regions of Tm 4-9 mutations....114
Figure D.1: Comparison of Tm 4-9 loop mutations.... 118
Figure E.1: NMR spectra of ice affinity purified protein.... 123

LIST OF SYMBOLS AND ABBREVIATIONS

AFGP	antifreeze glycoprotein
AFP	antifreeze protein
BMRB	BioMagResBank
CD	circular dichroism
CP/MAS	cross-polarization/magic-angle spinning
CSD	chemical shift deviation
$\Delta\delta/\Delta T$	temperature gradient
Dc	*Dendroides canadensis*
DQF-COSY	double quantum filtered correlated spectroscopy
DSS	2,2-dimethyl-2-silapentane-5-sulfonic acid
DTT	dithiothreitol
η	viscosity
HNHA	^{15}N-edited ^{1}H^{N}-^{1}H$^{\alpha}$ correlation experiment
HNHB	^{15}N-edited ^{1}H^{N}-^{1}H$^{\beta}$ correlation experiment
HPLC	high performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IAP	ice affinity purification
$^{3}J_{\alpha\beta}$	3 bond scalar coupling constant between spins Hα and Hβ
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	nuclear Overhauser effect spectroscopy
O.D.	outer diameter
PDB	Protein Data Bank
R_1	longitudinal relaxation rate
R_2	transverse relaxation rate
rmsd	root-mean-square deviation
S^2	generalized order parameter
sbw	spruce budworm
τ_c (τ_m)	overall rotational correlation time
T_1	longitudinal relaxation time
T_2	transverse relaxation time
TH	thermal hysteresis
THP	thermal hysteresis protein
Tm	*Tenebrio molitor*
TOCSY	total correlation spectroscopy

CHAPTER 1: General Introduction

1.1 Biological Significance

Water profoundly influences molecular interactions in biological systems (*1*). The ability of organisms to use or exclude water at both the macroscopic and microscopic level when necessary is fundamental for such functions as cell viability, utilization of single water molecules in enzyme catalyzed reactions, or exclusion of water in protein folding via the hydrophobic effect.

Ice poses a different set of issues for organism survival at cold temperatures, where exposure can be lethal. An example of this occurred in dramatic fashion this past April when approximately 200 000 kg of eastern cod washed ashore on the coast of Newfoundland (*2*). The recovered fish had ice crystals around their organs. Reports indicate that the water temperatures were the coldest measured in decades and the likely explanation is that they simply froze to death in water that was too cold for them.

Since most living organisms are killed when their tissues freeze, they must find ways to adapt. For many organisms, this involves behavioural adaptations such as migration to warmer climates or entrance into a hibernation state. However many species spend the winter, or in some cases their whole lives, in habitats which provide minimal protection from freezing temperatures. Yet they can endure exposure to temperatures well below the freezing point of their bodily fluids. In general, two cold survival strategies have developed: freeze tolerance and freeze avoidance (*3*, *4*).

Freeze tolerance is more complicated and therefore less common than freeze avoidance, but has developed in a number of varied species. Freeze tolerant animals, such as wood frogs, have developed quite an incredible strategy for surviving the winter, spending days or weeks with up to 65% of the total water in their body converted to ice (*5*). In this state, only the intracellular water remains liquid, while ice growth occurs all through the abdominal cavity, the lumen of the blood vessels, the ventricles of the brain and even grows in sheets between the skeletal muscle and the skin (*6*). These frozen animals show no signs of life, yet resume normal functions within minutes after thawing. Survival of this type requires careful regulation to prevent ice crystals from damaging tissues, to reduce metabolic stress and help tissues remain alive, and to produce high levels of cryoprotectants such as sugars and polyols to minimize osmotic effects (*6*).

Freeze avoiding animals, on the other hand, are able to maintain the liquid state of their bodily fluids at temperatures well below the usual minima for their habitat. They do this by keeping their body fluids in a supercooled state; that is, at a point below the equilibrium freezing point, using some very interesting adaptations (*7,8*). First, contact with potential nucleators is minimized to prevent seeding of ice formation at temperatures at or below the freezing point. Second, many animals produce high levels of sugar alcohols such as glycerol in their body fluids, which depress the freezing point and crystallization temperatures by colligative means. Most interestingly, some organisms also produce antifreeze proteins (AFPs) which specifically bind ice crystals and prevent them from growing to a size that could cause physical damage.

When AFPs were discovered over thirty years ago in the blood of Antarctic fishes (*9*), it was observed that the blood would freeze at –0.7 °C owing to the colligative effects of plasma solutes as mentioned above. These effects do not depend on the chemical nature of the molecules, but rather are proportional to the concentration of the solute (*10*). Seawater, with its high salt concentration, freezes at –1.9 °C, which means that marine fishes will typically encounter water that is more than a degree colder than their freezing point (*11*) and would freeze on contact with ice. To avoid this fate, AFPs are produced which bind to the surface of a seed ice crystal and control its growth (*12*). The adsorption of the AFP on the ice surface then restricts the addition of water to the ice crystal to the spaces between the bound AFPs, creating a localized curvature in the ice front. This makes addition of water to the ice lattice thermodynamically unfavourable and results in non-equilibrium freezing point depression without a corresponding change in the melting point (*13*). This difference between the freezing and melting temperatures is called thermal hysteresis (TH). Within this temperature gap, ice crystals are relatively static, remaining the same size for days (*14*) (Figure 1.1). This property of AFPs is used to detect and quantify AFP activity.

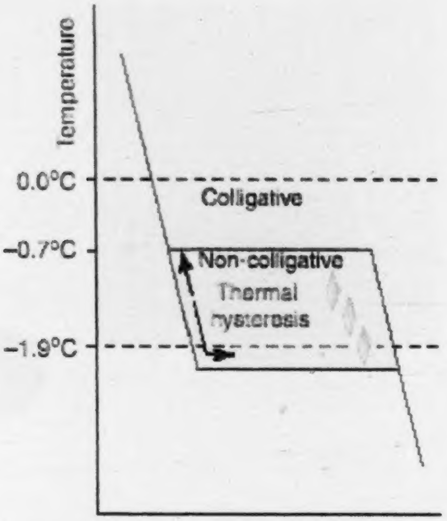


Figure 1.1: Fish achieve protection from freezing using a combination of colligative and non-colligative mechanisms. Seawater (blue) freezes at –1.9 °C, indicated by the transition to white. Fish blood (pink) freezes just below –1.9 °C, indicated by the lower red horizontal line. The total freezing point depression of the blood arises from a combination of the two components: colligative depression owing to blood solutes from 0 °C to –0.7 °C, and the non-colligative effects of the AFP binding to ice crystals. (Adapted from ref. *11*).

1.2 Structural Diversity of AFPs

AFPs have been isolated from various distinct groups of organisms, including fish (*7*), insects (*8*), plants, (*15,16*), and microorganisms (*17*). The fish AFPs have been classified into five types [antifreeze glycoprotein (AFGP) and AFP types I – IV] based on primary sequence and three-dimensional structural information (Table 1.1). Two types have thus far been characterized in insects, one from beetles (*Tenebrio molitor*, Tm, and *Dendroides canadensis*, Dc) and the other from the spruce budworm (sbw). These proteins are all vastly different in primary, secondary and tertiary structure (Table 1.1 and Figure 1.2) and yet all are able to bind to ice and shape its growth.

Table 1.1: Antifreeze proteins from fish and insects

Type	Size (kDa)	Repeat	Classification[a]
Fish			
AFGP	3-24	3 aa (AAT)	glycoprotein (O-glycosylated on T)
I	3-5	11 aa	alanine-rich α helix
II	14-24	none	C-type lectin fold, mixed α/β structure
III	7	none (dimer observed)	globular with short β strands
IV	12	22 aa[b]	predicted helical bundle
Insects			
TmAFP	8-10	12 aa (contains TCT)	right-handed β-helix
sbwAFP	9-12	15 aa (contains TXT)	left-handed β-helix

[a] see Figure 1.2 [b] repeat pattern unconfirmed, based on prediction

In fish, the AFGPs are formed from a tripeptide repeat of Ala-Ala-Thr, in which the threonine residues are O-linked to β-D-galactosyl-(1→3)-α-N-acetyl-D-galactosamine disaccharide moieties (*18*). The most often studied Type I AFP from winter flounder is a single α-helix with an 11 amino acid sequence repeat in which every third turn of helix bears a threonine residue (*19*). Sea raven Type II AFP has a mixture of α and β structure with a fold that is homologous to the carbohydrate-recognition domain of the Ca^{2+}-dependent lectin family of proteins (*20*). Type III AFP from eel pout is a small, globular β-sheet protein with a unique fold (*21*). Neither Type II nor Type III AFP from fish display any repeated sequence or regular placement of surface exposed residues. However, a natural dimer has been observed in Antarctic eel pout (*22*) and was recently characterized as having enhanced activity (*23*). The Type IV AFP from longhorn sculpin has not had its structure solved, but has been modeled as a helical bundle based on its high degree of sequence homology with apolipoproteins (*24*).

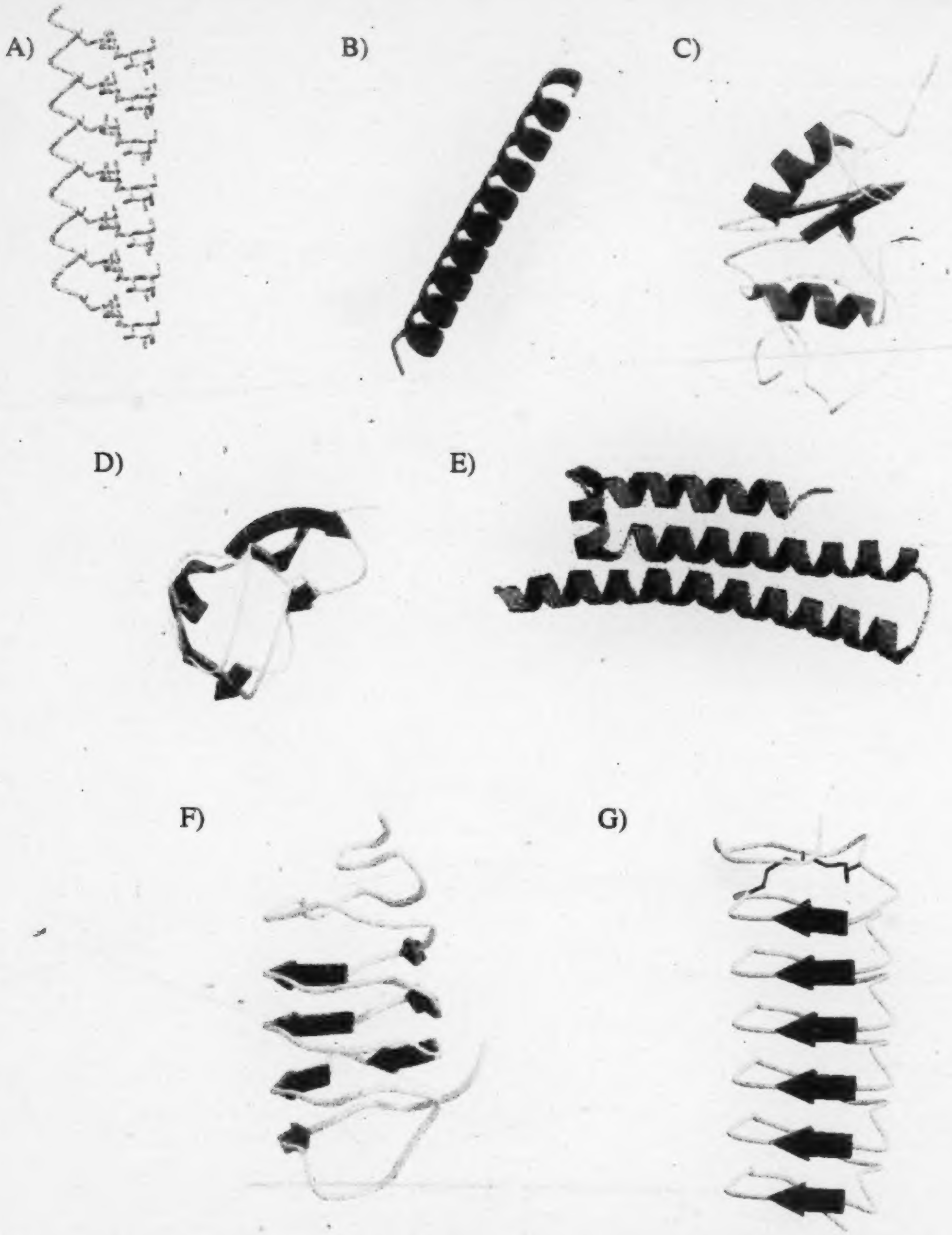


Figure 1.2: AFP structural diversity. Representative structures are drawn to scale for the five fish AFP types (A-E) and two insect AFP types (F, G), with α-helices in red, β-strands in blue, and coil in yellow (PDB accession codes are given in parentheses). A: AFGP drawn in extended form, with disaccharides shown. B: Type I AFP (1WFA). C: Type II AFP (2AFP). D: Type III AFP (1MSI). E: Type IV AFP modeled as a helical bundle. F: Spruce budworm AFP (1EWW). G: *Tenebrio molitor* AFP (1EZG). Figures generated using Molscript (*25*) and Raster3D (*26*).

Thermal hysteresis activity in insects was initially noted more than thirty years ago (*27*), however progress in characterizing their structures and activities proceeded much more slowly than for fish AFPs. Most of these studies have only occurred within the last 7 years. Insect AFPs have been isolated from spruce budworm (*Choristoneura fumiferana, 28*), the yellow mealworm beetle (*Tenebrio molitor, 29*), and the fire-coloured beetle (*Dendroides canadensis, 30*), which shares ~50% homology to the TmAFP. The AFPs from these two insect groups are both β-helical proteins, but are unrelated at the level of primary sequence and have different three-dimensional structures (Figure 1.2). Despite these differences, the ice-binding surfaces are remarkably similar (*31, 32*), consisting of two ranks of precisely aligned threonine residues. The insect AFPs have specific activities 10 – 100 times greater than the fish AFPs (*29*).

1.3 Ice Binding Mechanism

Ice binding is a specific interaction that occurs on defined ice planes. This has been determined by a method known as 'ice-etching' (*33*). In this technique, AFP in dilute solution is allowed to bind to an ice hemisphere where it adsorbs to its particular binding plane, but then is steadily overgrown. The binding surfaces are revealed by sublimation, which leaves behind a detectable AFP residue which can be interpreted in terms of the specific binding plane. This is corroborated by observations that AFP solutions shape the seed ice crystals during TH measurements (*34*). Furthermore, the TH activity is observed to be saturable, indicating a limited number of binding sites (*10*). Binding to a specific plane slows growth on that plane and a facet develops. As neighbouring ice fronts advance perpendicular to the inhibited face more of that plane is produced and therefore its growth is also inhibited. This eventually leads to complete stoppage of growth as only the inhibited plane is expressed.

All fish AFPs shape ice into hexagonal bipyramids by binding to pyramidal planes. The superior specific activity of insect AFPs has been attributed to their ability to bind to two planes of ice, as compared to just one for fish AFPs. This has been observed for sbwAFP, which has affinity for both the basal and primary prism planes, leading to a hexagonal plate shaped ice crystal (*32*). It has been suggested that the hexagonal bipyramidal ice crystal has weak points for containment at the two tips of the pyramids, which are the burst points for ice growth once the non-equilibrium freezing point is exceeded (*11*). The hexagonal plate shaped crystals of insect AFPs achieved by binding two planes reach much greater freezing point depressions before ice growth proceeds.

The use of TH measurements and ice crystal growth studies has led to a good understanding of AFP function at the macroscopic level. A description of the molecular mechanism however, has not been easily obtainable. One reason for this difficulty has already been introduced; there is a wide structural variation between different AFPs. Another factor is that the binding surface of the AFP-ice interaction cannot be directly examined. These points will be discussed in greater detail in the rest of this thesis.

An important feature of AFP-ice binding is that it must be considered irreversible, since any significant off-rate would lead to further ice growth (*35, 36*). In the general situation where an ice front is in equilibrium with an aqueous solution, slight undercooling of the solution below the equilibrium freezing point leads to water molecules joining the ice lattice, resulting in the advancement of the ice front. Most solutes, including nearly all proteins, are excluded from the ice and pushed ahead of the growing ice. For AFPs to fulfill their function of inhibiting ice growth by binding to specific planes, it has been suggested that the characteristic of adsorption necessary to prevent the migration of molecule with the advancing ice front is a perfectly matching interface (*36*). With a very tightly formed interface, water molecules cannot penetrate between the protein and the ice, leading to the observed stoppage of ice growth.

1.4 NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has been used to study biomolecules for many decades, but since the development of two-dimensional NMR methodology and strategies for sequence-specific resonance assignment (*37*), the field has undergone a dramatic revolution. NMR is now a mature field and a robust tool for protein structure determination. This is evidenced by the rapid growth in NMR structures deposited in the Protein Data Bank (PDB) – from 23 in 1990, to 2704 (or approximately 15% of the PDB holdings) as of August 5, 2003 (*38*).

Well-established ^{1}H homonuclear experimental protocols allow for routine structure determination for proteins in the 10 kDa range. With uniform isotope labeling using ^{15}N and ^{13}C in bacterial expression systems and a standard suite of heteronuclear NMR experiments (*39*), this limit increases to 25-30 kDa. Fractional incorporation of deuterium at certain side chain positions using specialized labeling schemes has increased the size limit even further (*40*). These techniques in combination with ever-improving experimental methods and instrumentation have allowed for resonance assignments to be completed for very large proteins (80 kDa; *41, 42*) and complexes (110 kDa; *43*). These extremes may not be readily achievable for many proteins, however the techniques being developed will surely impact the field greatly. It remains that for proteins up to 25 kDa, NMR is a particularly useful and well-developed method for three-dimensional structure determination.

The advantage of NMR is that it is a unique and powerful tool for studying protein dynamics in solution. The use of these techniques has grown tremendously in the last decade, especially as it relates to the study of protein function (*44*). Proteins are dynamic molecules that often undergo changes in conformation while performing their biological functions, such as ligand binding or enzyme reaction. The intrinsic dynamic properties of proteins themselves can provide information about the location and process of the conformational change and have been the focus of numerous biophysical studies. The utility of NMR in this regard arises from the ability to obtain site-specific dynamic information from nuclei at sites all over the protein (*45*). Furthermore, the dynamics that can be measured are not only due to structural heterogeneity, but relate to particular time-scales of motion that are biologically relevant. Protein internal motions span a wide range of amplitudes and rates, from fast (subnanosecond) bond vector fluctuations, to intermediate and exchange motions relating to local stability (milliseconds to microseconds), and slow movement of protein domains on the order of seconds to days (*46*). All of these time-scales are accessible using NMR; slow movements are reported using rates of magnetization transfer between protons with differing chemical shifts and proton exchange rates, while measurement of nuclear spin relaxation rates reports the faster internal motions, as well as the overall rotational diffusion of the molecule.

Recently, the relationship between local flexibility and conformational entropy has been addressed (*47*). Studies suggest that changes in conformational entropy make significant contributions to the free energy of protein binding to its target (*48, 49*). The conclusions reached indicate that proteins have considerable residual entropy and that changes in the functional state are associated with changes in the magnitude and distribution of the protein residual entropy (*44*). Conformational entropy increases may contribute to the stabilization of complexes, whereas decreases have been associated with enthalpy-entropy compensation (*50*). It seems likely that the key determinants of enthalpy-entropy compensation in protein-protein interactions are the fast subnanosecond fluctuations of the sidechain (*47*).

1.5 Objectives

In this thesis, I have investigated the structure and dynamics of the antifreeze protein from the *T. molitor* beetle in an attempt to characterize it fully. At the time I started working on this protein, the fish AFPs were well-studied but the proposed ice-binding mechanism was being called into question. As well, the insect AFPs had only just been cloned and no structures were yet known for these. Antifreeze proteins are very amenable to NMR investigation owing to their small size and stable nature. We embarked on these studies in collaboration with Dr. Peter Davies' laboratory at Queen's University with no real idea of what the structure would reveal. Despite the challenges presented by the repetitive TmAFP sequence during my initial assignment of the protein, I determined the NMR solution structure as discussed in Chapter 2. At the same time, we began to consider the concept of dynamics in the context of AFP structure. Since the form of the three-dimensional structure does not appear to be an important criterion for antifreeze protein function, dynamics can provide insight into the unusual activity and function of these fascinating proteins. The characterization of ^{15}N backbone dynamics for TmAFP is also addressed in Chapter 2; this represents the first study of backbone dynamics of any AFP. The observation that the backbone was uniformly rigid led us to study the role of the side chains in the binding interface. Chapter 3 and Chapter 4 both examine the behaviour of the threonine side chains of TmAFP, using different experimental strategies. Rotation about the χ_1 angle of threonine is discussed using motionally averaged scalar couplings in Chapter 3 and natural abundance ^{13}C relaxation experiments in Chapter 4. This provides two very different, yet complementary views regarding the restricted motion of these side chains. Use of ^{13}C relaxation at natural abundance also allows for examination of another backbone bond vector to supplement the view of the backbone provided by ^{15}N relaxation. In Chapter 5, the temperature dependence of the chemical shifts of the backbone amide protons is examined in relation to the hydrogen bonding interactions of the protein. All of these contributions together constitute the most complete look at the dynamics of any antifreeze protein to date and provide solid experimental support to an ice-binding mechanism that relies on surface complementarity between the AFP and ice.

1.6 References

1. Voet, D. and Voet, J.G. (1995) *Biochemistry, 2^{nd} ed.* John Wiley, New York.

2. CBC News, April 7, 8, and 10, 2003 (http://www.cbc.ca/news/)

3. Storey, K.B. and Storey, J.M. (1989) In *Animal Adaptation to Cold* (Wang, L.C.H., ed.) pp. 51-82, Springer-Verlag, Heidelberg.

4. Zachariassen, K.E. (1985) *Physiol. Rev.* **65**, 799-832.

5. Storey, K.B. and Storey, J.M. (1992) *Ann. Rev. Physiol.* **54**, 619-637.

6. Storey, K.B. and Storey, J.M. (2001) In *Protein Adaptations and Signal Transduction* (Storey, K.B. and Storey, J.M., eds.) pp. 1-19. Elsevier, Amsterdam.

7. Fletcher, G.L., Hew, C.L., and Davies, P.L. (2001) *Ann. Rev. Physiol.* **63**, 359-390.

8. Duman, J.G. (2001) *Ann. Rev. Physiol.* **63**, 327-357.

9. DeVries, A.L. and Wohlschlag, D.E. (1969) *Science* **163**, 1073-1075.

10. Yeh, Y. and Feeney, R.E. (1996) *Chem. Rev.* **96**, 601-617.

11. Jia, Z. and Davies, P.L. (2002) *TIBS* **27**, 101-106.

12. Raymond, J.A. and DeVries, A.L. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2589-2593.

13. Wilson, P.W. (1993) *Cryo-Letters* **14**, 31-36.

14. Davies, P.L., Baardsnes, J., Kuiper, M.J., and Walker, V.K. (2002) *Phil. Trans. R. Soc. Lond. B* **357**, 927-935.

15. Worrall, D., Elias, L., Ashford, D., Smallwood, M., Sidebottom, C., Lillford, P., Telford, J., Holt, C., and Bowles, D. (1998) *Science* **282**, 115-117.

16. Sidebottom, C., Buckley, S., Pudney, P., Twigg, S., Jarman, C., Holt, C., Telford, J., McArthur, A., Worrall, D., Hubbard, R., and Lillford, P. (2000) *Nature* **406**, 249-251.

17. Yamashita, Y., Nakamura, N., Omiya, K., Nishikawa, J., Kawahara, H., and Obata, H. (2002) *Biosci. Biotechnol. Biochem.* **66**, 239-247.

18. DeVries, A.L. (1983) *Ann. Rev. Physiol.* **45**, 245-260.

19. Yang, D.S., Sax, M., Chakrabartty, A., and Hew, C.L. (1988) *Nature* **333**, 232-237.

20. Gronwald, W., Loewen, M.C., Lix, B., Daugulis, A.J., Sönnichsen, F.D., Davies, P.L., and Sykes, B.D. (1998) *Biochemistry* **37**, 4712-4721.

21. Sönnichsen, F.D., Sykes, B.D., Chao, H., and Davies, P.L. (1993) *Science* **259**, 1154-1157.

22. Wang, X., DeVries, A.L., and Cheng, C.H. (1995) *Biochim. Biophys. Acta* **1247**, 163-172.

23. Baardsnes, J., Kuiper, M.J., and Davies, P.L. (2003) *J. Biol. Chem.* (In press).

24. Deng, G. and Laursen, R.A. (1998) *Biochim. Biophys. Acta* **1388**, 305-314.

25. Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24**, 946-950.

26. Merritt, E.A. and Bacon, D.J. (1997) *Methods Enzymol.* **277**, 505-524.

27. Ramsey, J.A. (1968) *Phil. Trans. R. Soc. Lond. B Biol. Sci.* **248**, 279-314.

28. Tyshenko, M.G., Doucet, D., Davies, P.L., and Walker, V.K. (1997) *Nat. Biotechnol.* **15**, 887-890.

29. Graham, L.A., Liou, Y-C., Walker, V.K., and Davies, P.L. (1997) *Nature* **388**, 727-728.

30. Duman, J.G., Li, N., Verleye, D., Goetz, F.W., Wu, D.W., Andorfer, C.A., Benjamin, T., and Parmelee, D.C. (1998) *J. Comp. Physiol. B* **168**, 225-232.

31. Liou, Y-C., Tocilj, A., Davies, P.L., and Jia, Z. (2000) *Nature* **406**, 322-324.

32. Graether, S.P., Kuiper, M.J., Gagné, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. (2000) *Nature* **406**, 325-328.

33. Knight, C.A., Cheng, C.C., and DeVries, A.L. (1991) *Biophys. J.* **59**, 409-418.

34. Houston, Jr., M.E., Chao, H., Hodges, R.S., Sykes, B.D., Kay, C.M., Sönnichsen, F.D., Loewen, M.C., and Davies, P.L. (1998) *J. Biol. Chem.* **273**, 11714-11718.

35. Knight, C.A. and DeVries, A.L. (1994) *J. Cryst. Growth* **143**, 301-310.

36. Knight, C.A. and Wierzbicki, A. (2001) *Cryst. Growth Des.* **1**, 439-446.

37. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. John Wiley & Sons, New York.

38. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000) *Nucleic Acids Res.* **28**, 235-242.

39. Ferentz, A.E. and Wagner, G. (2000) *Quart. Rev. Biophys.* **33**, 29-65.

40. Gardner, K.H. and Kay, L.E. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27**, 357-406.

41. Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L.E. (2002) *J. Am. Chem. Soc.* **124**, 10025-10035.

42. Tugarinov, V. and Kay, L.E. (2003) *J. Am. Chem. Soc.* **125**, 5701-5706.

43. Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wüthrich, K. (2000) *J. Am. Chem. Soc.* **122**, 7543-7548.

44. Wand, A.J (2001) *Nature Struct. Biol.* **8**, 926-931.

45. Kay, L.E. (1998) *Nature Struct. Biol.* **5 Suppl**, 513-517.

46. Ishima, R. and Torchia, D.A. (2000) *Nature Struct. Biol.* **7**, 740-743.

47. Spyracopoulos, L. and Sykes, B.D. (2001) *Curr. Opin. Struct. Biol.* **11**, 555-559.

48. Zidek, L., Novotny, M.V., and Stone, M.J. (1999) *Nature Struct. Biol.* **6**, 1118-1121.

49. Lee, A.L., Kinnear, S.A., and Wand, A.J. (2000) *Nature Struct. Biol.* **7**, 72-77.

50. Stone, M.J. (2001) *Acc. Chem. Res.* **34**, 379-388.

CHAPTER 2: The Solution Structure and Backbone Dynamics of the *Tenebrio molitor* Antifreeze Protein*

2.1 Introduction

Various organisms, including many marine fishes (*1*), insects (*2*), and plants (*3.4*), use antifreeze proteins (AFPs), also known as thermal hysteresis proteins, as a freeze avoidance strategy. For example, the yellow mealworm beetle *Tenebrio molitor* and the spruce budworm (*Choristoneura fumiferana*) produce AFP in their hemolymph to survive overwintering in the larval stage (*5*,*6*). Since their discovery in fish and insects more than thirty years ago (*7*,*8*), the biophysical properties of these proteins have inspired widespread interest. Their protective effect stems from their ability to bind to ice and thereby inhibit ice crystal growth. This interaction at the ice surface leads to a lowering of the non-equilibrium freezing point below the melting point, termed thermal hysteresis (for reviews, see references 9, *10*).

Even though it is well accepted that AFPs bind to ice and inhibit its growth, the actual mechanism of this binding is not well understood. In the past decade, the structures of numerous AFPs belonging to five different types have been solved, and none of these have any sequence or structural similarity to each other. In fish, type I AFP from winter flounder (*11*) is a single, long α-helix; type II AFP from sea raven has a homologous fold to the carbohydrate-recognition domain of the Ca^{2+}-dependent lectin family of proteins (*12*); the eel pout type III AFP is a small β-sheet protein with a unique fold (*13*,*14*); while a model of a four-helix bundle structure for type IV AFP from the longhorn sculpin has been made based on its high degree of sequence homology with the exchangeable apolipoproteins (*15*). More recently, the structures of two β-helical insect AFPs (*Tenebrio molitor*, TmAFP, and spruce budworm, sbwAFP) were solved (*16*,*17*). These insect AFPs are much more potent than their fish counterparts. For type I, type III and sbwAFP (and by analogy, TmAFP) putative ice-binding faces have been proposed based on mutation studies (*17-23*). The only feature these possible sites appear to have in common is that they are quite flat surfaces, and, in the case of type I and the insect AFPs, they have evenly spaced ranks of threonine that appear to match the ice lattice (*16, 24*). For sbwAFP and TmAFP, the increased thermal hysteresis activity displayed by these proteins could be due to these proteins binding to two different planes of ice (*17*), a suggestion that has been made for the Type III antifreeze protein as well (*14*).

---

* A version of this chapter has been published. Daley, M.E., Spyracopoulos, L., Jia, Z., Davies, P.L., and Sykes, B.D. 2002. Structure and dynamics of a β-helical antifreeze protein. *Biochemistry*, 41: 5515-5525.

Although structure often leads to insight into protein function, in this case structures alone have not allowed elucidation of a molecular mechanism for antifreeze protein binding to ice. In attempts to understand this interaction more fully, we have moved to study protein dynamics in the context of structure. Protein dynamics studies have become an important tool in understanding the processes underlying molecular recognition and the relationship of structure and function. Using heteronuclear, multi-dimensional NMR spectroscopy it is possible to study the dynamics of individual nuclei in proteins in a site-specific manner. In particular, measurement of NMR relaxation of backbone amide ^{15}N nuclei provides detailed information regarding backbone motions of the protein (*25*). In many cases of protein-ligand complexes, it is the flexibility of the protein and the ligand, as well as the flexibility of the binding interface that determine the affinity of the interaction (*26*). In this case, the ligand in question is ice, and the interface cannot be fixed in solution, so we must use alternate methods to probe this binding interaction. Recently, Graether et al. (*27*) have shown that type I AFP remains structured and behaves predictably until the solution freezes at –3°C. Similar experiments with TmAFP (Daley et al., unpublished) show that this AFP also does not denature under supercooling conditions and suggests that the native folded state is the active form. This is in contrast to the conformation and dynamics study of an antifreeze glycoprotein that appears to contain no permanent secondary structure and is active in its extended form (*28*). To our knowledge, these are the only dynamics studies on antifreeze and related proteins. It seems clear from this comparison that antifreeze proteins and glycoproteins have fundamentally different modes of action and the latter will not be discussed further.

In this paper, we present the solution structure of the *Tenebrio molitor* antifreeze protein, along with its backbone NMR relaxation data. This is the first extensive study of backbone dynamics of any antifreeze protein. We have completed full ^{15}N relaxation analysis at two temperatures, 30°C and 5°C, in order to determine the type of dynamic changes that occur in the backbone conformation of TmAFP as the temperature is lowered towards its physiologically active level.

2.2 Experimental Procedures

2.2.1 Sample preparation

The expression and purification of both unlabeled and ^{15}N-labeled TmAFP was as previously described (*29*). The ^{15}N-TmAFP sample was prepared for NMR spectroscopy by dissolving lyophilized protein in 90% H_2O / 10% D_2O containing 0.1 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). The final protein concentration was approximately 0.4 mM, and the pH was adjusted to 5.5 with microliter aliquots of 100 mM NaOD or DCl as required. The unlabeled TmAFP sample was prepared by dissolving the lyophilized protein in either 90% H_2O / 10% D_2O or 100% D_2O as required. The final protein concentration was approximately 1 mM, with the pH adjusted to 5.5 and 0.1 mM DSS added.

2.2.2 Assignment

NMR spectra for ^{1}H and ^{15}N chemical shift assignment were acquired using a Varian Unity 600 MHz spectrometer equipped with a 5 mm triple resonance probe and *z*-axis pulsed field gradients. All spectra for assignment were collected at 30°C. The spectral widths for ^{1}HN and ^{15}N were 7000.35 Hz and 1897.89 Hz, respectively. Experiments used for assignment were: ^{15}N-HSQC with 448 (HN) and 128 (^{15}N) real data points acquired with a total of 256 transients per t_1 increment, ^{15}N-edited TOCSY with 420 (HN), 116 (H), and 30 (^{15}N) real data points acquired with a total of 36 transients per increment, HNHA with 448 (HN), 52 (H), and 52 (^{15}N) real data points acquired with a total of 40 transients per increment, and HNHB with 512 (HN), 96 (H), and 30 (^{15}N) real data points acquired with a total of 32 transients per increment. Sequential assignments were made using an ^{15}N-edited NOESY acquired on a Varian INOVA 800 MHz spectrometer equipped with a 5 mm triple resonance probe and *x*, *y*, and *z*-axis pulsed field gradients. The spectral widths for ^{1}HN, ^{1}H, and ^{15}N were 10000 Hz, 8200 Hz, and 2188.4 Hz, respectively. The numbers of real data points acquired were 640 (HN), 172 (H), and 32 (^{15}N) with a total of 16 transients per increment. The mixing time was 100 ms. Following sequential assignment, the two proline residues were assigned from the δCHs using a natural abundance ^{13}C-HSQC experiment with 448 (^{1}H) and 306 (^{13}C) real data points acquired with a total of 352 transients per increment. The lysine and glutamine sidechains were then fully assigned using 2D TOCSY and NOESY experiments with 2048 (^{1}H) and 256 (^{1}H) real data points with a total of 64 transients per increment. The three previous experiments were collected at 600 MHz (spectral widths were 7000.35 Hz for ^{1}H and 2000 Hz for ^{13}C) on the unlabeled TmAFP sample in D_2O. Spectra were processed with NMRPipe (*30*) and analyzed using NMRView (*31*) programs.

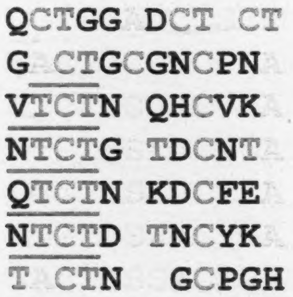


Figure 2.1: Amino acid sequence of the *T. molitor* AFP isoform used in this study. The repetitiveness of the sequence is highlighted by use of red for Cys, blue for Thr, light blue for Ser, and green for Ala. The underlined residues correspond to the regions of the protein that involve β-sheet secondary structure.

Owing to the repetitive sequence and structure of TmAFP, shown in Figure 2.1, the chemical shift assignment procedure used was atypical. Many residues are in similar environments in the protein and discrimination of chemical shifts was difficult. This was further complicated by the fact that NOEs between residues 12 amino acids apart in sequence were in many cases stronger than the sequential connections, which was misleading in the process of NOE assignment. These realizations were made once we analyzed the crystal structure of TmAFP (*16*). Using the PDB coordinates (prior to publication, 1EZG), we measured distances from the backbone amide proton to other protons in the protein and concluded that interloop distances were less than the sequential distances, since the secondary structure consists only of extended strands and turns. By thus correlating the interproton distances from the crystal structure with the NOE cross-peak intensities, a structure-specific sequential assignment of TmAFP was completed, with full ^{1}H and ^{15}N chemical shift assignments made. The crystal structure itself was not directly used in the assignment process.

2.2.3 Structure calculations

An ensemble of 50 TmAFP structures was generated from 672 distance restraints and 98 dihedral angle restraints using the simulated annealing program CNS version 1.0 (Crystallography & NMR System, *32*). Interproton distance restraints were obtained from the ^{15}N-edited NOESY with additional side chain NOE restraints obtained from the high-resolution homonuclear two-dimensional ^{1}H-^{1}H NOESY experiment. The ϕ dihedral angle restraints were obtained from an analysis of the $^3J_{HN\text{-}HA}$ coupling constants measured in the 3D HNHA experiment (*33*). The ψ dihedral angle restraints were determined by analysis of $d_{N\alpha}/d_{\alpha N}$ ratios; however, these were incorporated with typically large bounds (± 100°) and only in the regions of well-defined structure after initial structure calculations using NOE distance restraints (*34*). All structure calculations included the eight disulfide bonds, which were previously determined to have the following connectivities: Cys^2-Cys^{11}, Cys^8-Cys^{18}, Cys^{15}-Cys^{21}, Cys^{27}-Cys^{33}, Cys^{39}-Cys^{45}, Cys^{51}-Cys^{57}, Cys^{63}-Cys^{69}, Cys^{75}-Cys^{81} (*29*). The disulfide bonds were restrained to a distance of 2.02 ± 0.1 Å. No distance violations greater than 0.5 Å or dihedral angle violations greater than 5° were found. PROCHECK_NMR (*35*) was used to analyze the quality of the structures and VADAR (*36*) was used to assign secondary structure. Figures 2.3, 2.4, and 2.5 were generated using MOLSCRIPT (*37*) and Raster3D (*38*).

2.2.4 NMR spectroscopy for backbone dynamics

NMR spectra were acquired using Varian INOVA 500 MHz and Unity 600 MHz spectrometers equipped with 5 mm triple resonance probes and *z*-axis pulsed field gradients. ^{15}N-T_1, -T_2, and {^{1}H}-^{15}N NOE experiments were carried out at 500 and 600 MHz at both 30°C and 5°C using sensitivity-enhanced gradient HSQC pulse sequences developed by Farrow et al. (*39*). For measurement of ^{15}N-T_1 relaxation times at 30°C, delays of 11.1, 55.5, 122.1, 199.8, 277.5, 388.5, 499.5, 666, 888, and 1110 ms were used at 500 MHz and delays of 11.1, 55.5, 122.1, 199.8, 277.5, 388.5, 499.5, 666, 888, and 999 ms were used at 600 MHz. At 5°C, delays of 11.1, 55.5, 122.1, 199.8, 277.5, 388.5, 499.5, 610.5, 721.5, 832.5, 888, and 999 ms were used at 500 MHz and delays of 11.1, 55.5, 122.1, 199.8, 277.5, 388.5, 499.5, 610.5, 777, and 999 ms were used at 600 MHz. For measurement of ^{15}N-T_2 relaxation times at both 30°C and 5°C, delays of 16.61, 33.22, 49.83, 66.44, 83.05, 99.66, 116.27, 132.88, 149.49, and 166.1 ms were used at 500 MHz. Delays of 16.54, 33.09, 49.63, 66.18, 82.72, 99.26, 115.81, 132.35, 148.90, and 165.44 ms were used at 600 MHz. In order to obtain equilibrium, a 1.0 s delay between repetitions of the pulse sequence for ^{15}N-T_1 measurements was employed, while the delay for obtaining equilibrium during the measurement of ^{15}N-T_2 was 2.5 s. {^{1}H}-^{15}N steady-state NOEs were measured from two HSQC spectra acquired with (NOE experiment) and without (noNOE experiment) proton saturation prior to the first ^{15}N excitation pulse. Proton saturation at both 500 and 600 MHz was achieved using a train of 120° proton pulses with 5 ms pulse intervals for a total 3.5 s of saturation. Relaxation delays of 2 ms (NOE experiment) or 5 ms (noNOE experiment) between repetitions of the pulse sequence were used. The spectral widths for ^{1}H and ^{15}N were 5500.172 and 1670.007 Hz at 500 MHz and 7000.35 and 2000 Hz at 600 MHz, respectively. The number of real data points acquired for ^{1}H and ^{15}N, respectively, were 352 and 96 (30°C) or 128 (5°C) at 500 MHz and 436 and 96 (30°C) or 128 (5°C) at 600 MHz. A total of 32 or 64 transients per t_1 increment were collected for ^{15}N-T_1 and -T_2 at 30°C, while 48 or 52 transients were collected at 5°C. At 500 MHz, 112 (30°C) or 92 (5°C) transients were accumulated for measurement of {^{1}H}-^{15}N NOE, while at 600 MHz, 96 transients were collected at both 30°C and 5°C.

2.2.5 NMR data processing

All NMR data were processed with the NMRPipe software (*30*). Enhanced sensitivity data were processed using the ranceY.M macro. The F_1 (^{15}N) dimension was extended by

32 complex points using linear prediction before zero filling. The F_2 (^1H) dimension was multiplied by a 60°-shifted sine-bell function and the F_1 dimension was multiplied by a 75°-shifted squared sine-bell function before Fourier transformation. The F_1 and F_2 dimensions were baseline corrected by polynomial subtraction in the frequency domain. At 5°C, an additional window function was employed. The NMRView program (*31*) was used for peak picking of all {^1H-^{15}N}-HSQC spectra. The values of the peak intensities for the ^{15}N-T_1 and -T_2 measurements were fit to single-exponential, two-parameter decay curves using the xcrvfit program (software available at: http://www.pence.ca/ftp). Error in the ^{15}N-T_1 and -T_2 measurements was obtained from non-linear least-squares fits of the peak intensities to two-parameter exponential decays. Error in the {^1H}-^{15}N NOE values was calculated from baseplane noise values in ^{15}N-HSQC spectra acquired with and without proton saturation.

2.2.6 ^{15}N-Relaxation data analysis

In general, the dipolar interaction between the ^{15}N nucleus and its attached proton and the chemical shift anisotropy of the ^{15}N nucleus are the main contributions to the backbone amide T_1 and T_2 relaxation times and the ^{15}N-NOE enhancement (*40*). The theoretical expressions describing ^{15}N-T_1, -T_2, and {^1H}-^{15}N steady-state NOE relaxation parameters are given here as linear combinations of the appropriate spectral density functions:

$$\frac{1}{T_1} = D\left[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)\right] + C\left[J(\omega_N)\right] \quad (1)$$

$$\frac{1}{T_2} = \frac{D}{2}\left[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)\right] + \tfrac{1}{6}C\left[3J(\omega_N) + 4J(0)\right] \quad (2)$$

$$\mathrm{NOE} = 1 + (\gamma_H/\gamma_N)D\left[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)\right]T_1 \quad (3)$$

where $D = (\mu_0/4\pi)^2\left(\gamma_H{}^2\gamma_N{}^2\left(\frac{h}{2\pi}\right)^2/4r_{NH}{}^6\right)$ and $C = \Delta^2\omega_N{}^2/3$, ω_H and ω_N are the Larmor frequencies of ^1H and ^{15}N, respectively, γ_H is the magnetogyric ratio of ^1H (2.68×10^8 rad s^{-1} T^{-1}), γ_N is the magnetogyric ratio of ^{15}N (-2.71×10^7 rad s^{-1} T^{-1}), μ_0 is the permeability constant of free space ($4\pi \times 10^{-7}$ kg m s^{-2} A^{-2}), *h* is Planck's constant, r_{NH} is the nitrogen-proton internuclear distance (1.02 Å), and Δ is the difference between the parallel and perpendicular components of the ^{15}N chemical shift tensor (-160 ppm). The spectral density function is represented with the Lipari-Szabo model-free formalism (*41,42*), using a generalized order parameter S^2, which describes the degree of spatial restriction of each backbone amide HN–N bond vector. With the assumption of isotropic tumbling of the molecule, a single, overall rotational correlation time for the protein, and an associated internal correlation time, the spectral density function is defined as:

$$J(\omega)=\frac{2}{5}\left[\frac{S^2\tau_m}{\left(1+\omega^2\tau^2{}_m\right)}+\frac{\left(1-S^2\right)\tau}{\left(1+\omega^2\tau^2\right)}\right] \tag{4}$$

where $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$, with τ_m being the correlation time for overall molecular tumbling and τ_e the correlation time for internal motion. The relaxation data for each residue is then fit to a motional model describing the overall molecular tumbling of the protein, including one or two types of internal motions of various time-scales. For computational reasons the basic motional model is separated into five specific models, as previously described (*39*). These models are tested iteratively starting with the simplest model, and subsequently the more complex models, until the proposed model describes the measured relaxation times within 95% confidence limits. Model 1 uses a form of the function in equation 4, including the parameters S^2, describing the amplitude of the spatial restriction for each backbone amide vector varying from 0 (no spatial restriction) to 1 (complete spatial restriction), and τ_m. In model 1 it is assumed that τ_e is fixed at zero and does not contribute to relaxation. Model 2 includes S^2, τ_m, and τ_e, and it is assumed that the internal motions corresponding to τ_e are within the picosecond time-scale ($0 < \tau_e < \tau_m$) and contribute to the relaxation. Model 3 is a modification of model 1, which includes a parameter (R_{ex}, in s^{-1}) to describe the contribution of microsecond to millisecond time-scale internal motions to $1/T_2$. Model 4 is model 2 modified to include the R_{ex} parameter. Model 5 invokes internal motion occurring on two time-scales (*43,44*). It includes an order parameter (S^2_f) for picosecond time-scale internal motions, an order parameter (S^2_s) for nanosecond time-scale internal motions that are faster than τ_m but slower than τ_e, and a correlation time (τ_s) for nanosecond time-scale internal motions. The correlation time τ_e, for picosecond time-scale internal motions, is assumed not to contribute to the relaxation (as in model 1). The extended model spectral density function is given in equation 5:

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{\left(1+\omega^2\tau^2{}_m\right)} + \frac{\left(S^2{}_f - S^2\right)\tau'_s}{\left(1+\omega^2\tau'^2{}_s\right)}\right] \tag{5}$$

where $S^2 = S^2{}_f S^2{}_s$ and $\tau'_s = \tau_s\tau_m / \left(\tau_s + \tau_m\right)$. In order to determine the most appropriate model to describe each residue, the parameters for each spectral density function are adjusted to minimize the following χ^2 function:

$$\chi^2 = \frac{\left(T_{1c} - T_{1e}\right)^2}{\sigma^2{}_{T1}} + \frac{\left(T_{2c} - T_{2e}\right)^2}{\sigma^2{}_{T2}} + \frac{\left(\mathrm{NOE}_c + \mathrm{NOE}_e\right)^2}{\sigma^2{}_{\mathrm{NOE}}} \tag{6}$$

where the subscripts c and e indicate calculated and experimental values, respectively, and σ is the error of the individual relaxation parameters used.

In the case of rigid body anisotropic reorientation, the spectral density function has been fully detailed elsewhere (*45,46*); for example in equations 3 and 4 in the paper by Tjandra et al. (*46*). The six parameters (D_{xx}, D_{yy}, D_{zz}, θ, ϕ, and ψ) describing the orientation and amplitude of the principal components of the diffusion tensor in the molecular frame are optimized. The diffusion parameters are then extracted by minimizing the error function:

$$\chi^2 = \sum_n \left\{\left[\left(R_2^{\mathrm{meas}}/R_1^{\mathrm{meas}}\right) - \left(R_2^{\mathrm{calc}}/R_1^{\mathrm{calc}}\right)\right]_n / \sigma_n\right\}^2 \tag{7}$$

where σ is the uncertainty in the experimental R_2/R_1 ratio. The summation is performed over all residues defined as rigid using the following equations as proposed by Tjandra et al. (*46,47*):

$$\mathrm{NOE} < \overline{\mathrm{NOE}} - 1.5\ \mathrm{SD} \tag{8a}$$

$$\frac{\langle T_2\rangle - T_{2,n}}{\langle T_2\rangle} - \frac{\langle T_1\rangle - T_{1,n}}{\langle T_1\rangle} > 1.5\ \mathrm{SD} \tag{8b}$$

where the T_1 and T_2 averages are taken over residues that have not been excluded because of a low NOE (eq. 8(a)). SD is the standard deviation of the function calculated for these residues.

For the diffusion tensor and internal mobility analysis, the Cartesian coordinates of TmAFP were taken from the crystal structure. The fit of the components and orientation of the diffusion tensor is performed using an in-house written program (*48*) and visualized with the program TENSOR2 (*49*). The relaxation parameters are analyzed with respect to isotropic, axially symmetric, and fully asymmetric rotational tumbling models using TENSOR2. The significance of the improved fit of each model is tested to ensure that the reduction in χ^2 is not due only to the introduction of additional parameters into each subsequently more complex model.

The global isotropic correlation time has been optimized using those residues that do not display obvious signs of relaxation active mobility, using the exclusion criteria detailed in equations 8a and 8b above. With the assumption of negligible internal mobility and exchange, the T_1/T_2 ratios of these residues should be dependent only on the overall reorientation of the molecule.

With the orientation and the components of the diffusion tensor optimized using the selected vectors, the internal mobility, shown here for the case of axially symmetric rotational diffusion, is characterized using the spectral density function extended to include fast internal motions:

$$J(\omega) = \frac{2}{5} S^2{}_f \sum_{j=1}^{3} A_j \left\{ \frac{S^2{}_s \tau_j}{1 + \left(\omega\tau_j\right)^2} + \frac{(1 - S^2{}_s)\tau'_j}{1 + \left(\omega\tau'_j\right)^2} \right\} \tag{9}$$

with $\tau'_j = \tau_j \tau_s \big/ \left(\tau_j + \tau_s\right)$, $\tau_1^{-1} = 6D\perp$, $\tau_2^{-1} = 5D\perp + D\|$, $\tau_3^{-1} = 2D\perp + 4D\|$, $A_1 = \left(3\cos^2\theta - 1\right)^2 \big/ 4$, $A_2 = 3\sin^2\theta\cos^2\theta$, $A_3 = (3/4)\sin^4\theta$, where θ is the angle between the NH bond vector and the unique axis of the principal frame of the diffusion tensor. This is analogous to the isotropic Lipari-Szabo formalism, as the fast internal motion (τ_i << $(6D)^{-1}$) is assumed to be independent of the overall rotational tumbling. The statistical tests performed to determine the significance of the derived models use the same criteria for model selection as in the isotropic case described above.

2.3 Results

2.3.1 Solution structure of TmAFP

The recombinant TmAFP used in this study consists of 85 residues, with the N-terminal Met used to initiate translation in *E. coli* numbered 0. In this way sequence numbering remains consistent with the native protein. The extreme regularity of the sequence and structure as well as the abundance of small amino acid side chains (Figure 2.1) made it very difficult to assign the ^{1}H and ^{15}N chemical shifts of this protein. Despite this, almost complete ^{1}H and ^{15}N NMR chemical shift assignments of TmAFP were obtained using the structure-specific assignment procedure discussed in the experimental procedures.

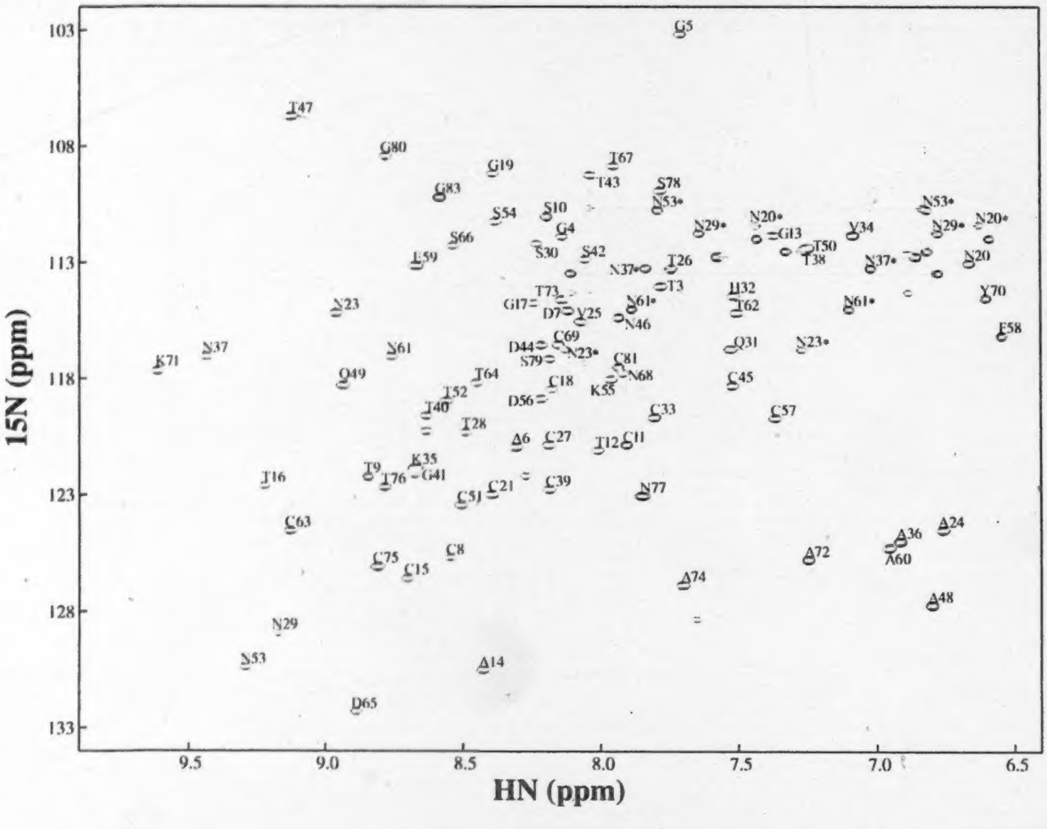


Figure 2.2: ^{1}H-^{15}N HSQC spectrum of TmAFP at 30°C, showing the complete backbone amide assignments and the wide chemical shift dispersion.

Backbone amide protons were unambiguously assigned for 79 out of the 83 non-proline residues at both 30°C (Figure 2.2) and 5°C (not shown). Resonances for the N-terminal

Met^0, Gln^1, and Cys^2, as well as the C-terminal His^{84} were not observed. As can be seen in the ^{1}H-^{15}N HSQC spectrum in Figure 2.2, the resonances are generally well-dispersed and the spectra are of high resolution. In addition, complete assignment of all side chains resonances was accomplished, with the exception of the four residues mentioned above which were not observed, and the side chain amide groups of three asparagines and the three glutamines that could not be unambiguously distinguished.

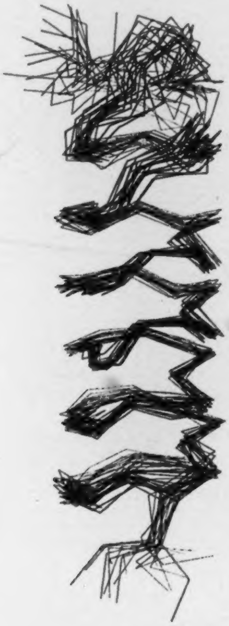


Figure 2.3: The tertiary structure of *Tenebrio molitor* antifreeze protein. Shown is the Cα trace of the superimposition (using residues 13-81) of the main chain atoms from the 20 lowest energy NMR-derived structures of TmAFP (PDB code 1L1I).

Restraints for 672 interproton distances (148 long-range, ($|i-j| \geq 5$); 40 medium-range, ($2 \leq |i-j| \leq 4$); 202 sequential ($|i-j| = 1$) and 282 intra-residue ($i = j$)) and 98 ϕ and ψ dihedral angles were derived as described in the experimental procedures. An ensemble of structures was generated using these restraints (Figure 2.3). The statistics for this ensemble are summarized in Table 2.1. Overall, the calculated structures exhibit good covalent geometry and agree well with the experimental data. The single structure closest to the mean of the ensemble was selected for illustrative purposes and the following discussion.

Table 2.1: Structural Statistics for the Ensemble of 20 refined structures of TmAFP

	Ensemble	Selected structure
r.m.s.d. from NOE distances (Å) (672)	0.022 ± 0.001	0.021
r.m.s.d. from dihedral angle restraints (degrees) (98)	0.37 ± 0.11	0.396
r.m.s.d. from idealized covalent geometry		
Bonds (Å)	0.0025 ± 0.0002	0.0022
Angles (degrees)	0.37 ± 0.01	0.35
Impropers (degrees)	0.23 ± 0.02	0.22
Cartesian coordinate r.m.s.d. (Å)		
Backbone atoms (N, Cα, C)	0.8 ± 0.1	0.6
All heavy atoms	1.1 ± 0.1	1.0
Ramachandran plot analysis		
Residues in most favored regions	53.3 %	52.8 %
Residues in additionally allowed regions	45.2 %	47.2 %
Residues in disallowed regions	1.5%	0%

TmAFP is a right-handed parallel β-helix consisting of seven repetitive 12-amino-acid loops (Figure 2.4). The consensus sequence for these repeated loops is xTCTxSxxCxxA, as described previously (*50*) and shown in Figure 2.1. The overall shape of the protein is a flattened cylinder approximately 32 Å long, with a 14 × 8 Å pseudo-rectangular cross section. The structure is constrained by a series of intraloop disulfide bonds running through the middle of the helix core. All 16 cysteines in TmAFP are involved in these disulfide bridges. Six of the eight disulfide bonds are spaced 6 residues apart and are aligned in the centre of each loop, except the N-terminal one (Figure 2.4). In the N-terminal region, the other two disulfide bonds (Cys^{2}-Cys^{11} and Cys^{8}-Cys^{18}) do not conform to this regular pattern; however, the N-terminal loop formation is not disrupted. The structure is further stabilized by a network of hydrogen bonds whose existence is supported by the observation of long-lived amide proton resonances in the two-dimensional TOCSY spectrum in D_2O. These hydrogen bonds are present in the β- and γ-turns of the individual helical loops, as well as in the stacking of adjacent loops (*16*). This is due to the extremely similar backbone conformations of each loop of the helix.

Owing to the tightness of this β-helix, with only 12 residues per loop, TmAFP has essentially no hydrophobic core and certainly no room inside the helix for long side chains. In addition to the disulfide-bonded cysteines that bisect each loop, the only other residues that fit inside the core are the conserved alanine and serine residues that line the right and left-hand sides of the protein, respectively, for the view in Figure 2.4C. All the rest of the side chains, including the relatively few hydrophobic ones, project out into solution.

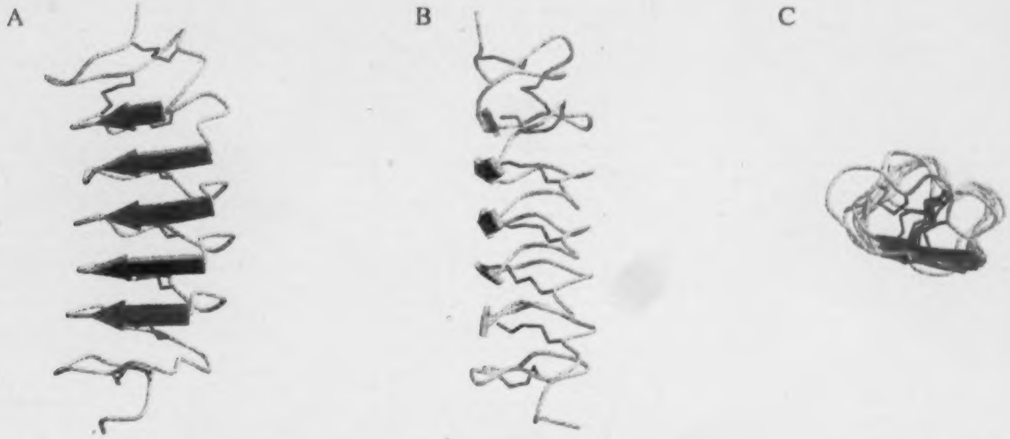


Figure 2.4: Ribbon representation of the TmAFP structure closest to the mean. The β-sheets, shown in green, are located at residues 14-16, 25-28, 37-40, 49-52, and 61-64. Disulfide bonds, shown in red, are located across each loop and between loops 1 and 2 at the N-terminus. (B) is rotated 90° to the left to emphasize the relative flatness of the proposed ice-binding face and (C) shows the view down the helix axis.

Each loop of the helix, except the N- and C-terminal ones, contributes a 3 - 4 residue β-strand to the sheet that is formed along one side of the protein. This sheet contains the array of threonine residues from the TCT motif that is the putative ice-binding face. The rest of the loop consists of extended coil structure as well as β- and γ-turns. The entire protein constitutes a well-defined structure, with r.m.s. distributions about the mean coordinate positions of 0.8 ± 0.1 Å for backbone (207) atoms and 1.1 ± 0.1 Å for all heavy (471) atoms from residues 13 - 81. The N-terminal 12 residues are poorly defined due to the smaller number of NOE restraints in this region.

This structure agrees very well with that previously determined by x-ray crystallography (*16*) (Figure 2.5). The r.m.s.d. between the representative solution structure and the x-ray structure is 1.3 Å for backbone (207) atoms and 2.0 Å for all heavy (471) atoms of the six well-defined loops as described above. The main difference is again in the N-terminal region, which is loosely defined in the solution structure. When the four C-terminal loops (residues 37-81) are used for superimposition, the r.m.s.d. between the two structures for the six well-defined loops drops to 1.1 Å for the backbone (207) atoms; however, when the three N-terminal loops (from residues 10-37) are superimposed, the backbone r.m.s.d. is increased to 1.6 Å.

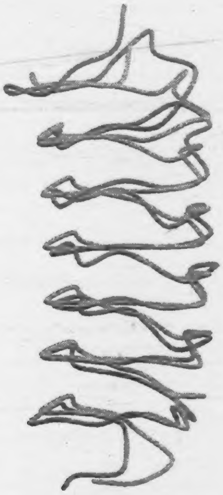


Figure 2.5: NMR and x-ray crystallographic structures of TmAFP. Shown are superimposed backbone traces of TmAFP solution (blue) and crystal (red) structures. The superimposition is for residues 13-81 with an r.m.s.d. of 1.35 Å. The view is the same as in Figure 2.4A.

2.3.2 ^{15}N-T_1, -T_2, and NOE data

The backbone dynamics of TmAFP have been studied at two temperatures, 30°C and 5°C. At 30°C, 73 residues at 500 MHz and 75 residues at 600 MHz were used for the relaxation analysis, due to {^{1}H-^{15}N}-HSQC spectral overlap. At 5°C, 72 residues at 500 MHz and 73 residues at 600 MHz were used.

The values of ^{15}N-T_1, -T_2, and {^{1}H}-^{15}N NOE as a function of residue number are shown for 30°C in Figure 2.6 and for 5°C in Figure 2.7. The profiles of ^{15}N-T_1, -T_2 and {^{1}H}-^{15}N NOE values for the two resonance frequencies display similar patterns as a function of residue number. The regular periodic pattern of the T_1, T_2, and T_1/T_2 values is obvious,

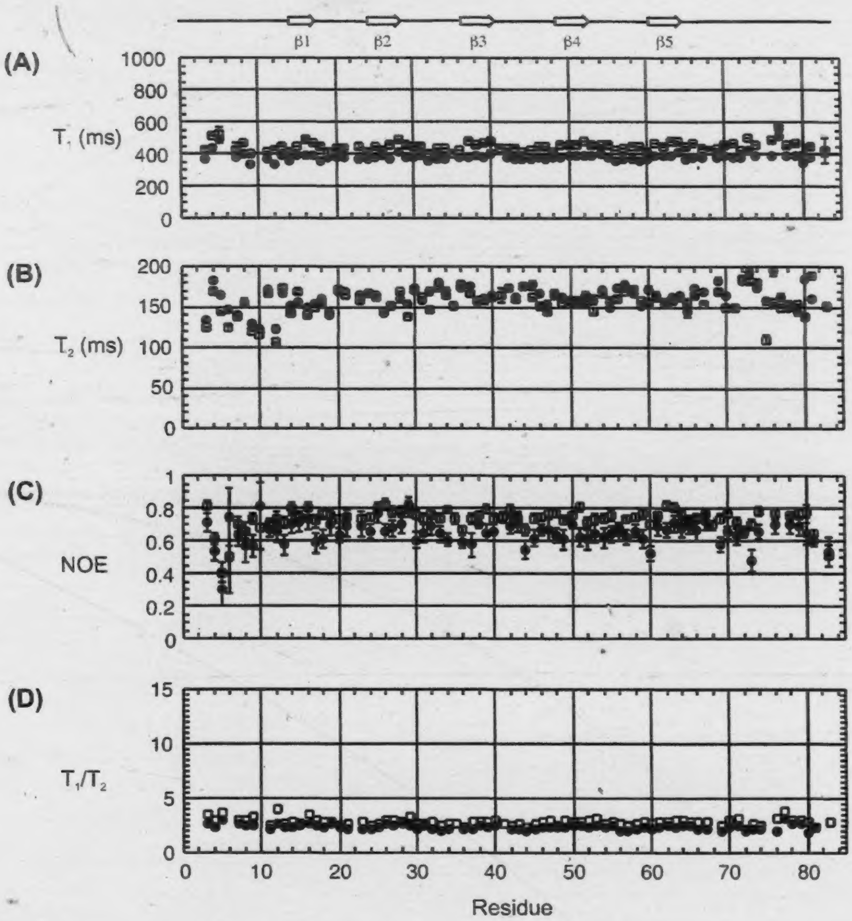


Figure 2.6: Plots of (A) ^{15}N-T_1, (B) ^{15}N-T_2, (C) {^{1}H}-^{15}N NOE, and (D) T_1/T_2 ratios at 30°C with individual error bars plotted for 500 (•) and 600 (□) MHz. Average T_1/T_2 ratios calculated for residues with negligible internal motion (NOE > 0.6 (500 MHz) or NOE > 0.65 (600 MHz) were 2.36 ± 0.21 for T_1^{500}/T_2^{500} (•) and 2.84 ± 0.29 for T_1^{600}/T_2^{600} (□). A schematic diagram of secondary structure is drawn above the panel with arrows depicting the β-strands. The residues in structured regions are: β1 (14-16), β2 (25-28), β3 (37-40), β4 (49-52), and β5 (61-64).

corresponding to the 12-amino-acid repeat sequence and loop structure. For proteins with $\tau_m \geq 5$ ns (ie. in the slow tumbling limit), ^{15}N-T_1 values are largely dependent on the rates of motion occurring at the ^{15}N Larmor frequency, and therefore exhibit characteristic magnetic field strength dependence (Figure 2.6). At 30°C the T_1^{500} is 381 ± 22 ms with an average error of 7 ms, and T_1^{600} is 445 ± 27 ms with an average error of 7

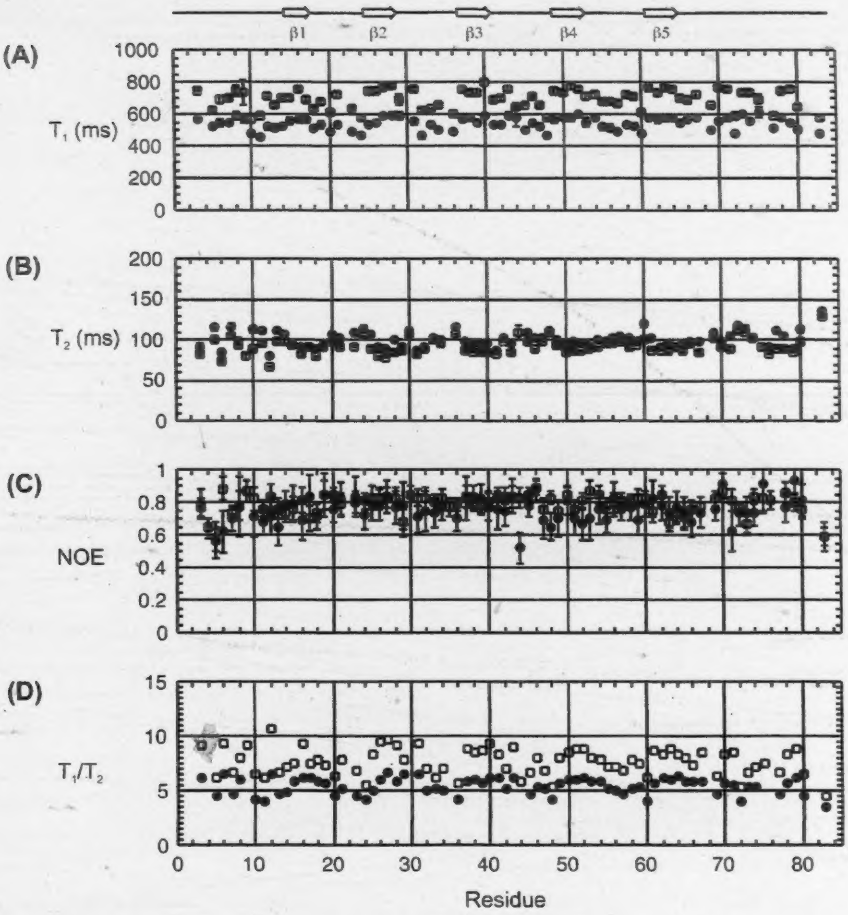


Figure 2.7: Plots of (A) ^{15}N-T$_1$, (B) ^{15}N-T$_2$, (C) {^{1}H}-^{15}N NOE, and (D) T$_1$/T$_2$ ratios at 5°C with individual error bars plotted for 500 (•) and 600 (□) MHz. Average T$_1$/T$_2$ ratios calculated for residues with negligible internal motion (NOE > 0.6 (500 MHz) or NOE > 0.65 (600 MHz) were 5.46 ± 0.72 for T_1^{500}/T_2^{500} (•) and 7.75 ± 1.18 for T_1^{600}/T_2^{600} (□).

ms for all residues. The magnetic field strength dependence becomes more obvious at 5°C (Figure 2.7) where the T_1^{500} is 542 ± 39 ms with an average error of 14 ms, and T_1^{600} is 695 ± 59 ms with an average error of 14 ms over all residues. ^{15}N-T$_2$ values in the slow tumbling limit are largely determined by the zero frequency value of the spectral density, J(0), and are therefore expected to be similar at the different magnetic field strengths, as long as there are no chemical or conformational exchange contributions to T$_2$. At 30°C, the average T_2^{500} is 162 ± 14 ms with an average error of 3 ms, and T_2^{600} is 156 ± 16 ms with an average error of 2 ms. At 5°C these values drop as expected due to slower tumbling and the average T_2^{500} is 101 ± 10 ms with an average error of 2 ms, and T_2^{600} is 92 ± 10 ms with an average error of 2 ms over all residues. $\{^1H\}$-^{15}N NOE measurements are particularly sensitive to fast internal motions on the picosecond time scale ($0 < \tau_e < \tau_m$), and again a similar pattern is observed at the two magnetic field strengths as a function of residue number. At 30°C the average NOE^{500} is 0.65 ± 0.07 with an average error of 0.06 and NOE^{600} is 0.74 ± 0.05 with an average error of 0.04 for all residues. At 5°C, the pattern holds and NOE^{500} is 0.75 ± 0.08 with an average error of 0.10 and NOE^{600} is 0.80 ± 0.06 with an average error of 0.06 over all residues.

2.3.3 Rotational diffusion anisotropy and overall correlation time

To calculate the global τ_m, the first step is to remove residues with NOE values below a certain threshold. This is done because the NOE is sensitive to internal dynamics on the picosecond-nanosecond time-scale, and those residues with motions that are faster than τ_m on this time-scale must be eliminated prior to the calculation. An NOE cutoff value of 0.6 was used at 500 MHz and an NOE value of 0.65 was used at 600 MHz. A total of 56 out of 73 residues (30°C) and 68 out of 72 residues (5°C) from the 500 MHz data and 70 out of 75 residues (30°C) and 70 out of 73 residues (5°C) from the 600 MHz data set were selected by the NOE criteria for use in the global τ_m calculation. Residues that were removed from the data sets were typically from the N- and C-terminal regions of the protein.

Those residues which did not have their T_1/T_2 ratios excluded due to the NOE criteria described above were used to estimate the global τ_m. Residues with a T_1/T_2 ratio above or below one standard deviation of the mean for all residues were also removed (*46, 47*). The relaxation parameters for each magnetic field strength were analyzed individually. At 30°C, the average T_1/T_2 ratios of the selected residues are 2.36 ± 0.21 and 2.84 ± 0.29 at 500 MHz and 600 MHz, respectively, while at 5°C the T_1/T_2 ratios are 5.46 ± 0.72 and 7.75 ± 1.18 at 500 MHz and 600 MHz. A total of 42 residues (30°C) or 41 residues (5°C) were used to calculate the global τ_m at 500 MHz, while a total of 53 residues (30°C) or 48 residues (5°C) were used at 600 MHz. τ_m was calculated on a per residue basis using the spectral density model 2 (eq.4) for isotropic rotational diffusion and the results were averaged to give the global τ_m. Global τ_ms of 4.2 ns (500 MHz) and 4.0 ns (600 MHz) were obtained at 30°C. At 5°C, a global τ_m of 7.7 ns was obtained at 500 MHz and 7.9 ns at 600 MHz.

From the observation of the periodic pattern of the raw dynamic data, anisotropy in the rotational tumbling of TmAFP is postulated. The crystal structure of TmAFP was used to calculate the normalized lengths of the principal axes of the inertia tensor, and the ratio was 1.00 : 0.52 : 0.45. Thus, anisotropy in the tumbling is indeed expected. The rotational diffusion anisotropy was examined for all residues not affected by large-amplitude fast picosecond and slower millisecond time-scale motions, in a similar manner to the procedure described by Pawley et al. (*51*). The relaxation parameters T_1, T_2, and NOE were analyzed with respect to isotropic, axially symmetric and fully asymmetric rotational tumbling models using a grid search to find the minimum in the squared difference measured and calculated T_1/T_2 ratios using an in-house written program (*48*). All residues for which relaxation data were available were considered for the anisotropic rotational tumbling analysis, unless they were shown to be subject to significant internal motions, as determined by the filtering method above. The rotational tumbling of TmAFP is best characterized using the axially symmetric model with $D_{\|}/D_{\perp} \cong$ 1.45 at 30°C and $D_{\|}/D_{\perp} \cong$ 1.65 at 5°C for both magnetic field strengths (Table 2.2). A statistical *F*-test, shown in Table 2.2, validated the inclusion of additional parameters in the axially symmetric rotational diffusion model, while the fully asymmetric model was over-parametrized (*52*). For the axially symmetric model, the probability that the improvement in fit by the inclusion of additional parameters occurred by chance is 0 at 5°C for both magnetic fields, and 3.2×10^{-11} (500 MHz) or 2.2×10^{-13} (600 MHz) at 30°C. The axially symmetric rotational diffusion tensor was found to be prolate in shape ($D_{\|}/D_{\perp} > 1$).

The orientation of the diffusion tensor with respect to the molecular structure of TmAFP is shown in Figure 2.8. D_z (or $D_{\|}$), the unique axis of the rotational diffusion tensor, is parallel to the long axis of the β-helix. The orientation of the amide bond vectors of the backbone varies depending on which of the 12 positions in the loop it occupies. As will be expanded upon in the next section, this analysis was carried out in order to prevent the misinterpretation of relaxation parameters by the assumption of isotropic rotational diffusion in the internal mobility (model-free) analysis.

Table 2.2: Rotational Diffusion Parameters by Isotropic, Axially Symmetric, and Asymmetric Models at 500 and 600 MHz and Two Temperatures (30°C and 5°C) for TmAFP[a]

model/ field/temp	τ_m[b](ns)	$D_\parallel/D_\perp$[c]	D_x/D_y	θ[d]	ϕ[d]	φ[d]	E[e]	E_v[f]	F_x[g]
isotropic[h]									
500 (30°C)	4.2	1	1				709.4	11.4	
600 (30°C)	4.0	1	1				2274.4	35.5	
500 (5°C)	7.7	1	1				1442.2	22.9	
600 (5°C)	7.9	1	1				2584.1	38.6	
ax. sym[h]									
500 (30°C)	3.9	1.43	1	61.2	-133.6		292.1	4.9	28.1
600 (30°C)	3.8	1.50	1	62.8	-130.8		813.8	13.3	36.5
500 (5°C)	7.1	1.65	1	60.2	-120.6		194.9	3.2	128.0
600 (5°C)	7.3	1.64	1	58.5	-121.0		474.3	7.4	94.9
asym[h]									
500 (30°C)	4.0	1.38	0.87	62.2	-132.1	-24.2	282.1	4.9	1.0
600 (30°C)	3.8	1.51	0.94	62.4	-131.7	-112.9	808.4	13.7	0.2
500 (5°C)	7.1	1.67	0.94	60.7	-121.3	-44.4	193.1	3.3	0.3
600 (5°C)	7.3	1.64	0.99	58.4	-121.0	0.1	474.1	7.6	0.01

[a] Values were calculated using the crystal structure of TmAFP (PDB entry 1EZG). [b] $\tau_m=1/6D$. [c] $D_\parallel/D_\perp = 2D_z/(D_x+D_y)$. [d] Euler angles describing the orientation of the components of the diffusion tensors, D_z (θ and ϕ) and $D_{x,y}$ (φ). [e] Error function $E = \sum_N \left(T_{1,e}/T_{2,e} - T_{1,c}/T_{2,c}\right)^2 / \sigma^2_{T_1/T_2}$, where N is the number of residues used in the fit. [f] Reduced error function $E_v=E/(N-m)$, where N and m are the number of residues and variables used in the fit, respectively. [g] $F_x=(E_m-E_{m-x})/xE_{v,m-x}$, where E and E_v are defined above and x is the number of additional variables in the fit. Larger F_x values justify the use of additional variables (Gagne et al., 1998). [h] For the isotropic, axially symmetric, and fully asymmetric anisotropic analyses, 63 (30°C) and 64 (5°C) residues at 500 MHz and 65 (30°C) and 68 (5°C) residues at 600 MHz were used.

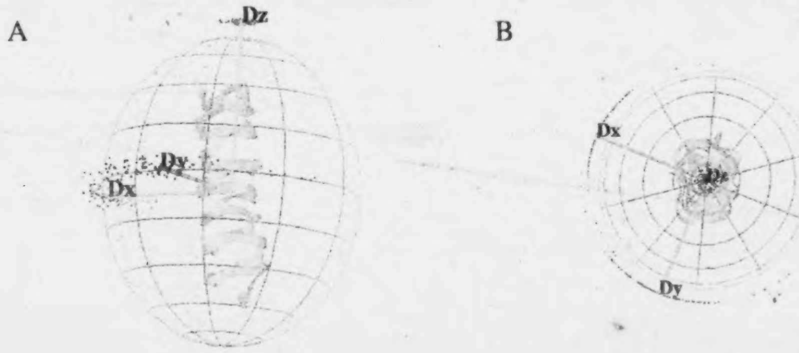


Figure 2.8: Rotational diffusion tensor of TmAFP. The best fit rotational diffusion tensor of TmAFP is shown superimposed on the TmAFP structure closest to the mean. The tensor is visualized as a three-dimensional ellipsoid with its axes marked. The orientation of the molecule in (A) is the same as in Figure 2.4B, while (B) shows the view down the helix axis from the position in (A).

2.3.4 Model-free analysis and internal mobility

The Lipari-Szabo model-free approach (*41,42*) is widely used to interpret backbone amide ^{15}N relaxation parameters. The experimental ^{15}N-T_1, -T_2, and {^{1}H}-^{15}N NOE relaxation parameters were fit to spectral density models 1-5. The spectral density parameters from each data set (500 and 600 MHz, at both 30°C and 5°C) were optimized individually (Figure 2.9) to allow assessment of the consistency between magnetic field strengths. This is possible to do since the motions that give rise to NMR relaxation are not dependent on magnetic field strengths, except for a slight alignment at higher magnetic field strengths (*53*).

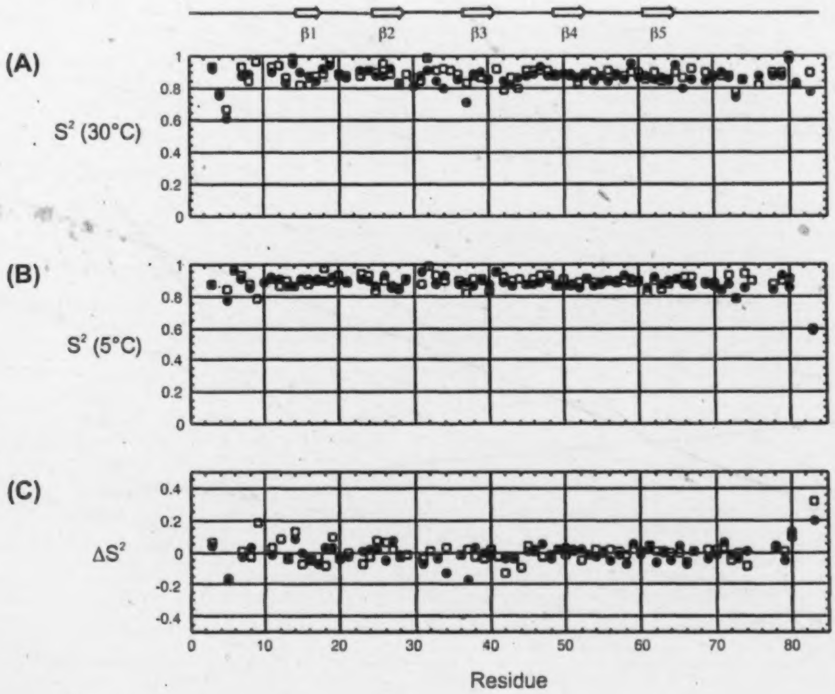


Figure 2.9: Plots of (A), order parameters (S^2) at 30°C, (B) S^2 at 5°C, and (C) $\Delta S^2 = S^2(30°C) - S^2(5°C)$ for 500 (•) and 600 (□) MHz data.

The S^2 values obtained agree very well between the two magnetic field strengths, as expected. At 30°C, the average S^2 values for the whole protein backbone are 0.87 at 500 MHz and 0.88 at 600 MHz (Figure 2.9A). These S^2 values are typical of well-defined secondary structure in proteins. At 5°C, the average S^2 values are almost identical, 0.88

at 500 MHz and 0.89 at 600 MHz (Figure 2.9B). Only some residues in the N-terminal loop and the very C-terminal residues display any motion of varying amplitudes and rates. For TmAFP, both the β-sheet and extended loop regions of the protein are highly structured.

All of the model-free analysis was performed accounting for rotational diffusion anisotropy (eq.9), as the failure to include this will lead to misidentification of τ_e and/or R_{ex} terms as genuine slow motions and conformational exchange (*46*). As well, S^2 values calculated assuming isotropic rotational diffusion are also affected and appear artificially lowered, implying the molecule is more flexible and undergoes more internal motion than is actually occurring. For TmAFP, as shown in Figure 2.8, D_z ($D_{\|}$) is parallel to the long axis of the β-helix. Around each loop of the helix, the orientations of the N-NH bond vector in each of the 12 positions vary from parallel to the helical axis to nearly perpendicular. These different orientations with respect to the unique axis of the diffusion tensor are what give rise to the periodic pattern displayed by the experimental relaxation data.

2.4 Discussion

The study of protein backbone dynamics provides detailed and site-specific information about the conformational changes that can occur upon ligand binding. Nuclear magnetic resonance spectroscopy is sensitive to a wide range of motions of various time scales, and is particularly applicable to this type of dynamic study. For TmAFP, the proposed ice-binding region is the array of threonine residues that form the parallel β-sheet on one face of the protein (*16*). The parameters extracted from these ^{15}N NMR relaxation studies show that TmAFP has similar restricted internal mobility throughout the protein backbone, at both temperatures studied. Therefore, both the ice-binding and opposite faces of TmAFP are formed by well-defined and rigid structural elements and the overall backbone dynamics are not affected as TmAFP approaches its physiologically relevant temperature.

This observed backbone rigidity of the ligand binding site is a recurring theme in several recent publications (*54,55*), in contrast to observations that ligand binding is often related to increased active site flexibility (*56*). In other recent studies, the heteronuclear NMR spin relaxation data has been interpreted in terms of its contributions to protein entropy and heat capacity (*57*). In the case of TmAFP, the ligand it binds is ice, which has an extremely regular spacing of its repetitive elements. If lattice matching to this array is to occur, it follows that the binding site in question would be rigid to ensure the correct positioning of the required protein constituents in close proximity. For TmAFP we propose that the hydrophobic, rigid binding site reduces the entropic penalty for the free energy of binding the protein to ice.

An examination of the structure of TmAFP suggests that a β-helix provides an ideal scaffold to produce the required rigidity of the ice-binding site. Although parallel β-helix structures have been discovered in bacteria and fungi, including other right-handed β-helices such as pectate lyase (*58*), rhamnogalacturonase A (*59*) and the P22 tailspike protein (*60*), and the left-handed β-helix domain of N-acetylglucosamine acyltransferase (*61*), all these proteins are much larger and less strictly regular than TmAFP. They typically contain 22 to 27 residues per β-helical loop and all have normal hydrophobic cores. Since no dynamic studies on these larger β-helices have been undertaken, we cannot state whether these β-helices will also demonstrate a similar degree of rigidity, although it is certainly a distinct possibility. Nonetheless, the compact size and shape of the TmAFP β-helix as compared to other β-helical structures, coupled with the extensive disulfide- and hydrogen-bonding, structures the helical loops to such a high degree that even the lack of a hydrophobic core does not affect its stability. It is interesting to note that another insect AFP, that of spruce budworm, has utilized a β-helical structure with a triangular architecture as a scaffold for antifreeze activity (*17*), while β-helical structures have also been postulated for the leucine-rich carrot AFP (*3*) and recently characterized ryegrass AFP (*62*). However, a β-helical AFP structure has not yet been seen in fishes.

In conclusion, we have determined that the entire TmAFP molecule is a highly structured and rigid protein. This rigidity appears to be inherent in the β-helical structure and does not appear to be a consequence of the lowering of the solution temperature towards physiologically active temperatures. From this observation, we suggest that the β-helix structure is an ideal scaffold for antifreeze protein activity and the key element to its success is its repetitive structure and associated rigidity, which allows it to match the solid ice crystal lattice as it binds. Given the wide variety of observed AFP structures in fish, it would be informative to study backbone dynamics of these AFPs to compare their proposed ice-binding face dynamics. As well, since it is clear that side chains play very important roles in binding interfaces, a detailed look at their dynamic behavior will be the next logical step in the attempt to elucidate the molecular mechanism of antifreeze protein binding to ice.

2.5 References

1. Fletcher, G.L., Hew, C.L., and Davies, P.L. (2001) *Annu. Rev. Physiol.* **63**, 359-390.

2. Duman, J.G. (2001) *Annu. Rev. Physiol.* **63**, 327-357.

3. Worrall, D., Elias, L., Ashford, D., Smallwood, M., Sidebottom, C., Lillford, P., Telford, J., Holt, C., and Bowles, D. (1998) *Science* **282**, 115-117.

4. Sidebottom, C., Buckley, S., Pudney, P., Twigg, S., Jarman, C., Holt, C., Telford, J., McArthur, A., Worrall, D., Hubbard, R., and Lillford, P. (2000) *Nature* **406**, 249-251.

5. Graham, L.A., Liou, Y-C., Walker, V.K., and Davies, P.L. (1997) *Nature* **388**, 727-728.

6. Tyshenko, M.G., Doucet, D., Davies, P.L., and Walker, V.K. (1997) *Nat. Biotechnol.* **15**, 887-890.

7. DeVries, A.L., and Wohlschlag, D.E. (1969) *Science* **163**, 1073-1075.

8. Grimstone, A.V., Mullinger, A.M., and Ramsay, J.A. (1968) *Philos. Trans. R. Soc. London B* **253**, 343-382.

9. Davies, P.L., and Sykes, B.D. (1997) *Curr. Opin. Struct. Biol.* **7**, 828-834.

10. Yeh, Y., and Feeney, R.E. (1996) *Chem. Rev.* **96**, 601-617.

11. Yang, D.S., Sax, M., Chakrabartty, A., and Hew, C.L. (1988) *Nature* **333**, 232-237.

12. Gronwald, W., Loewen, M.C., Lix, B., Daugulis, A.J., Sönnichsen, F.D., Davies, P.L., and Sykes, B.D. (1998) *Biochemistry* **37**, 4712-4721.

13. Sönnichsen, F.D., DeLuca, C.I., Davies, P.L., and Sykes, B.D. (1996) *Structure* **4**, 1325-1337.

14. Jia, Z., DeLuca, C.I., Chao, H., and Davies, P.L. (1996) *Nature* **384**, 285-288.

15. Deng, G., Andrews, D.W., and Laursen, R.A. (1997) *FEBS Lett.* **402**, 17-20.

16. Liou, Y-C., Tocilj, A., Davies, P.L., and Jia, Z. (2000) *Nature* **406**, 322-324.

17. Graether, S.P., Kuiper, M.J., Gagné, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. (2000) *Nature* **406**, 325-328.

18. Chao, H., Houston Jr., M.E., Hodges, R.S., Kay, C.M., Sykes, B.D., Loewen, M.C., Davies, P.L., and Sönnichsen, F.D. (1997) *Biochemistry* **36**, 14652-14660.

19. Haymet, A.D.J., Ward, L.G., Harding, M.M., and Knight, C.A. (1998) *FEBS Lett.* **430**, 301-306.

20. Zhang, W., and Laursen, R.A. (1998) *J. Biol. Chem.* **273**, 34806-34812.

21. Haymet, A.D.J., Ward, L.G., and Harding, M.M. (1999) *J. Am. Chem. Soc.* **121**, 941-948.

22. Baardsnes, J., Kondejewski, L.H., Hodges, R.S., Chao, H., Kay, C.M., and Davies, P.L. (1999) *FEBS Lett.* **463**, 87-91.

23. Chao, H., Sönnichsen, F.D., DeLuca, C.I., Sykes, B.D., and Davies, P.L. (1994) *Protein Sci.* **3**, 1760-1769.

24. Knight, C.A., Cheng, C.C., and DeVries, A.L. (1991) *Biophys. J.* **59**, 409-418.

25. Kay, L.E. (1998) *Nat. Struct. Biol.* **5 Suppl**, 513-517.

26. Ishima, R., and Torchia, D.A. (2000) *Nat. Struct. Biol.* **7**, 740-743.

27. Graether, S.P., Slupsky, C.M., Davies, P.L., and Sykes, B.D. (2001) *Biophys. J.* **81**, 1677-1683.

28. Lane, A.N., Hays, L.M., Tsvetkova, N., Feeney, R.E., Crowe, L.M., and Crowe, J.H. (2000) *Biophys. J.* **78**, 3195-3207.

29. Liou, Y-C., Daley, M.E., Graham, L.A., Kay, C.M., Walker, V.K., Sykes, B.D., and Davies, P.L. (2000) *Protein Expr. Purif.* **19**, 148-157.

30. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR* **6**, 277-293.

31. Johnson, B.A., and Blevins, R.A. (1994) *J. Biomol. NMR* **4**, 603-614.

32. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J-S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) *Acta. Cryst. D* **54**, 905-921.

33. Vuister, G.W., and Bax, A. (1993) *J. Am. Chem. Soc.* **115**, 7772-7777.

34. Gagné, S.M., Tsuda, S., Li, M.X., Chandra, M., Smillie, L.B., and Sykes, B.D. (1994) *Protein Sci.* **3**, 1961-1974.

35. Laskowski, R.A., Rullmann, J.A.C., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996) *J. Biomol. NMR* **8**, 477-486.

36. Willard, L., Wishart, D.S., and Sykes, B.D. (1997) *VADAR Version 1.2*, University of Alberta, Edmonton, Alberta, Canada.

37. Kraulis, P.J. (1991) *J. Appl. Cryst.* **24**, 946-950.

38. Merritt, E.A., and Bacon, D.J. (1997) *Methods in Enzymology* **277**, 505-524.

39. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) *Biochemistry* **33**, 5984-6003.

40. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.

41. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4546-4559.

42. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc.* **104**, 4559-4570.

43. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.* **112**, 4989-4991.

44. Clore, G.M., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. (1990) *Biochemistry* **29**, 7387-7401.

45. Woessner, D.E. (1962) *J. Chem. Phys.* **37**, 647-654.

46. Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. (1995) *J. Am. Chem. Soc.* **117**, 12562-12566.

47. Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) *J. Biomol. NMR* **8**, 273-284.

48. Gagné, S.M., Tsuda, S., Spyracopoulos, L., Kay, L.E., and Sykes, B.D. (1998) *J. Mol. Biol.* **278**, 667-686.

49. Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. (2000) *J. Biomol. NMR* **16**, 23-28.

50. Liou, Y-C., Thibault, P., Walker, V.K., Davies, P.L., and Graham, L.A. (1999) *Biochemistry* **38**, 11415-11424.

51. Pawley, N.H., Wang, C., Koide, S., and Nicholson, L.K. (2001) *J. Biomol. NMR.* **20**, 149-165.

52. Bevington, P.R., and Robinson, D.K. (1992) *Data Reduction and Error Analysis for the Physical Sciences*, 2nd ed., McGraw-Hill, New York.

53. Tjandra, N., Grzesiek, S., and Bax, A. (1996) *J. Am. Chem. Soc.* **118**, 6264-6272.

54. Suh, J-Y., Spyracopoulos, L., Keizer, D.W., Irvin, R.T., and Sykes, B.D. (2001) *Biochemistry* **40**, 3985-3995.

55. Poget, S.F., Freund, S.M.V., Howard, M.J., and Bycroft, M. (2001) *Biochemistry* **40**, 10966-10972.

56. Forman-Kay, J.D. (1999) *Nature Struct. Biol.* **6**, 1086-1087.

57. Spyracopoulos, L., and Sykes, B.D. (2001) *Curr. Opin. Struct. Biol.* **11**, 555-559.

58. Yoder, M.D., Keen, N.T., and Jurnak, F. (1993) *Science* **260**, 1503-1507.

59. Petersen, T.N., Kauppinen, S., and Larsen, S. (1997) *Structure* **5**, 533-544.

60. Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R., and Reinemer, P. (1994) *Science* **265**, 383-386.

61. Raetz, C.R., and Roderick, S.L. (1995) *Science* **270**, 997-1000.

62. Kuiper, M.J., Davies, P.L., and Walker, V.K. (2001) *Biophys. J.* **81**, 3560-3565.

CHAPTER 3: The Role of Side Chain Conformational Flexibility in Surface Recognition by *Tenebrio molitor* Antifreeze Protein[*]

3.1 Introduction

Antifreeze proteins (AFPs) present in various organisms, including fish (*1*), insects (*2*), and plants (*3*, *4*), help them to survive the freezing conditions of the environments in which they live. The yellow mealworm beetle *Tenebrio molitor*, for example, produces AFP (abbreviated TmAFP) in its hemolymph while overwintering in the larval stage. AFPs bind to the ice surface, inhibiting ice crystal growth and depressing the freezing point of the solution below the melting point. The separation of the non-equilibrium freezing temperature and the melting temperature in a solution containing ice crystals is referred to as thermal hysteresis (for a review, see ref *5*). At millimolar concentrations, the TmAFP can account for 5.5 °C of thermal hysteresis (*6*). This is much more active (10-100 X) than the well-studied fish AFPs (*7*).

Binding to the ice surface is believed to proceed via an adsorption-inhibition mechanism, although details of this interaction are not understood at a molecular level (*8*). AFP binding to ice is specific and occurs on defined planes. This is supported by observations that AFP solutions are able to shape the ice crystals to produce particular ice crystal morphologies (*9*). Crystal growth is slowed or stopped completely on the plane to which the AFP binds. For example, fish AFPs shape the ice crystal into a hexagonal bipyramid by binding to the pyramidal plane, while insect AFPs can bind to both prism and basal planes, shaping the ice crystal into a hexagonal plate (*10*). Ice etching studies can reveal the specific ice surface bound by an AFP and also support this hypothesis (*11*).

Early attempts to define the nature of the AFP-ice interaction focused on hydrogen bonding of polar side chains to the ice lattice, either through the inclusion of hydrogen bonding side chains into the ice lattice (lattice occupancy) (*12*) or by a hydrogen bond match with ice surface oxygen atoms (lattice matching) (*13*). The models were based on the crystal structure of the α-helical Type I AFP from winter flounder, the best characterized and simplest AFP to study. It is composed of 37 amino acids, with a tandemly repeated 11-amino acid unit with consensus sequence TX_2N/DX_7, where X is generally alanine (*14*). The 16.5 Å spacing of the *i*, *i*+11 threonine residues matches the 16.7 Å distance of the water molecules on the pyramidal plane. However, further experimentation has led to diminished importance for the hydrogen bonding interaction since mutation of threonine to serine, preserving the hydroxyl group and therefore the hydrogen bonding ability of the side chain, caused a severe loss in activity, while mutation to valine, preserving the methyl group and the shape of the side chain, resulted in a relatively minor loss of activity (*15-18*). Additionally, a redefinition of the ice-binding site of this AFP has occurred on the basis of mutation studies in which the alanine 17 to leucine mutant abolished antifreeze activity (*19*). On the basis of these studies and a recent examination of Type III AFP ice-binding (*20*), a universal ice-binding mechanism relying on hydrophobic and van der Waals interactions is emerging.

---

[*] A version of this chapter has been published. Daley, M.E. and Sykes, B.D. 2003. The role of side chain conformational flexibility in surface recognition by *Tenebrio molitor* antifreeze protein. *Protein Science*, 12: 1323-1331.

The mechanism of ice-binding has also been complicated by questions concerning the relative rigidity or flexibility of the side chains. Upon refinement of the Type I crystal structure, the ice-binding threonine residues were observed to all adopt a rigid side chain conformation with a χ_1 of -60° (*13*). These authors suggested that the rigidity and common orientation of these side chains are critical for the ice-binding mechanism. However, solution NMR studies at low temperatures indicated that the threonine residues of Type I AFP were in fact flexible, and can sample many possible rotameric states prior to ice binding (*21*).

The highly active TmAFP is an ideal candidate for this type of mechanistic study. TmAFP is a small (8.4 kDa) highly disulfide-bonded, right-handed parallel β-helix consisting of seven tandemly repeated 12 amino acid loops. Its crystal structure revealed an array of threonine residues on the β-sheet side of the protein that all adopted the same χ_1=-60° rotameric conformation, and the spacing of the hydroxyl groups is a near-perfect match to the prism plane of ice and approximates the spacing on the basal plane (*22*). In addition, the crystal structure also contained bound external water molecules that, along with the threonine hydroxyls, mimic a section of the ice lattice; the first time that this has been observed in an AFP structure. Furthermore, this threonine array has been defined as the ice-binding face by extensive mutagenesis (*23*). The threonines not on the ice-binding face serve as internal controls. We have previously solved the NMR solution structure of TmAFP and performed analysis of the ^{15}N backbone relaxation parameters which revealed it to be a well-folded and rigid protein with restricted backbone internal mobility throughout, at both 30 and 5 °C (*24*). In this study, we have investigated the orientations of the threonine side chains in solution to examine the role of flexibility of these side chains in this binding interface using high-resolution two-dimensional DQF-COSY experiments to measure the $^{3}J_{\alpha\beta}$ coupling constants. These coupling constants, in combination with NOE data, were used to determine the conformational states of the threonine side chains. These observations are important to the consideration of surface residues, both for understanding antifreeze protein activity, as well for other systems involved in molecular recognition.

3.2 Results

The amino acid sequence and structure of TmAFP are shown in Figure 3.1A and 3.1B, respectively. The threonine residues arrayed on the β-sheet side of the protein, which have been identified by mutagenesis as the ice-binding residues (*23*), are highlighted in blue. Those threonine residues that are not part of the ice-binding face are coloured red. Threonine 9, which is also not involved in ice binding, is coloured green since coupling constant data could not be obtained for this residue.

A

QCTGGADCTSCT
GACTGCGNCPNA
VTCTNSQHCVKA
NTCTGSTDCNTA
QTCTNSKDCFEA
NTCTDSTNCYKA
TACTNSSGCPGH

B

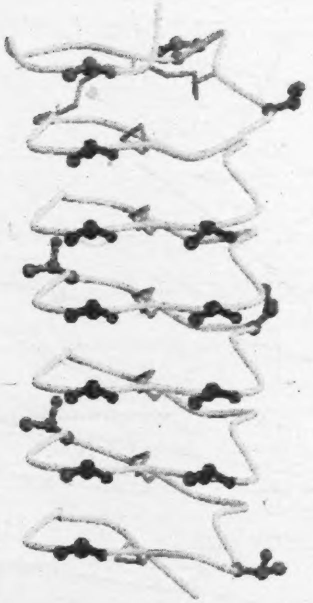


Figure 3.1: Sequence and structure of the *T. molitor* AFP isoform used in this study. (A) Amino acid sequence. The 17 threonine residues are coloured to emphasize their positions in the protein. The threonines in blue are those that form the ice-binding face, the threonines in red and green are not part of the ice-binding face. Coupling constant data could not be obtained for T9 (in green). (B) Structure of TmAFP (PDB 1EZG) with threonine side chains displayed in ball-and-stick representation and coloured as for the sequence in (A). Figure generated using Molscript (*25*) and Raster3D (*26*).

To ensure there was no change in the structure of TmAFP in normal versus supercooled water, the ^{1}H NMR spectra of TmAFP in D_2O as a function of temperature were examined (Figure 3.2). For clarity, only the amide region from 6 to 10 ppm is displayed,

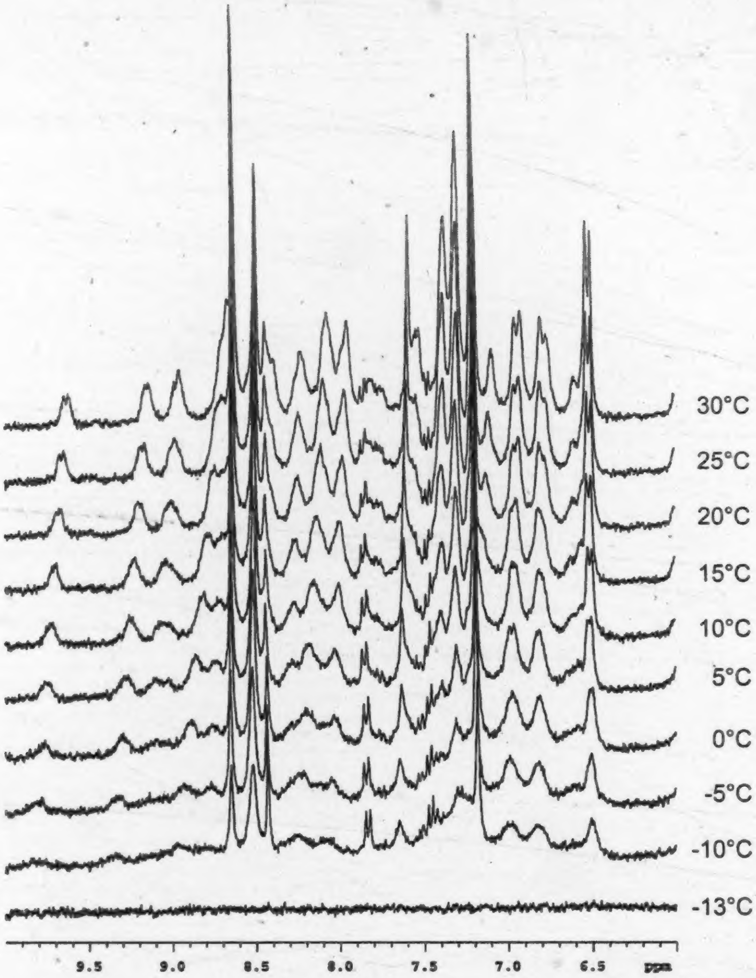


Figure 3.2: ^{1}H NMR spectra at 300 MHz of TmAFP in 100% D_2O. Samples were cooled from 30 °C to -20 °C as outlined in Materials and Methods.

corresponding to the HN resonances of buried residues. This region is a particularly sensitive indicator of protein folding and has been successfully used to assess TmAFP folding (*27*). As the temperature is decreased, there are no changes in chemical shift, only a line-broadening which is attributable to the increase in solvent viscosity and does not represent a structural change at low temperature. This has been extensively studied using Type I AFP with similar results (*28*). As shown in the final spectrum in Figure 3.2, the sample eventually froze at -13 °C at which point liquid state NMR experiments could no longer be conducted.

Characterization of the threonine ^{1}H-^{1}H $^3J_{\alpha\beta}$ coupling constants of TmAFP began with the identification of the threonine Hα-Hβ crosspeaks in the DQF-COSY spectrum (Figure 3.3). The advantages of DQF-COSY are that double quantum filtration is used to eliminate singlet resonances to help simplify the spectrum and resolution is improved since both the diagonal and crosspeaks are in pure absorption phase. The disadvantage is that it is less sensitive by a factor of two than conventional COSY. This was compensated for by making a sample with as high a protein concentration as possible and by collecting the data at the highest available field. Chemical shifts for this protein have previously been reported (*24*) and are deposited in the BMRB (accession number 5323).

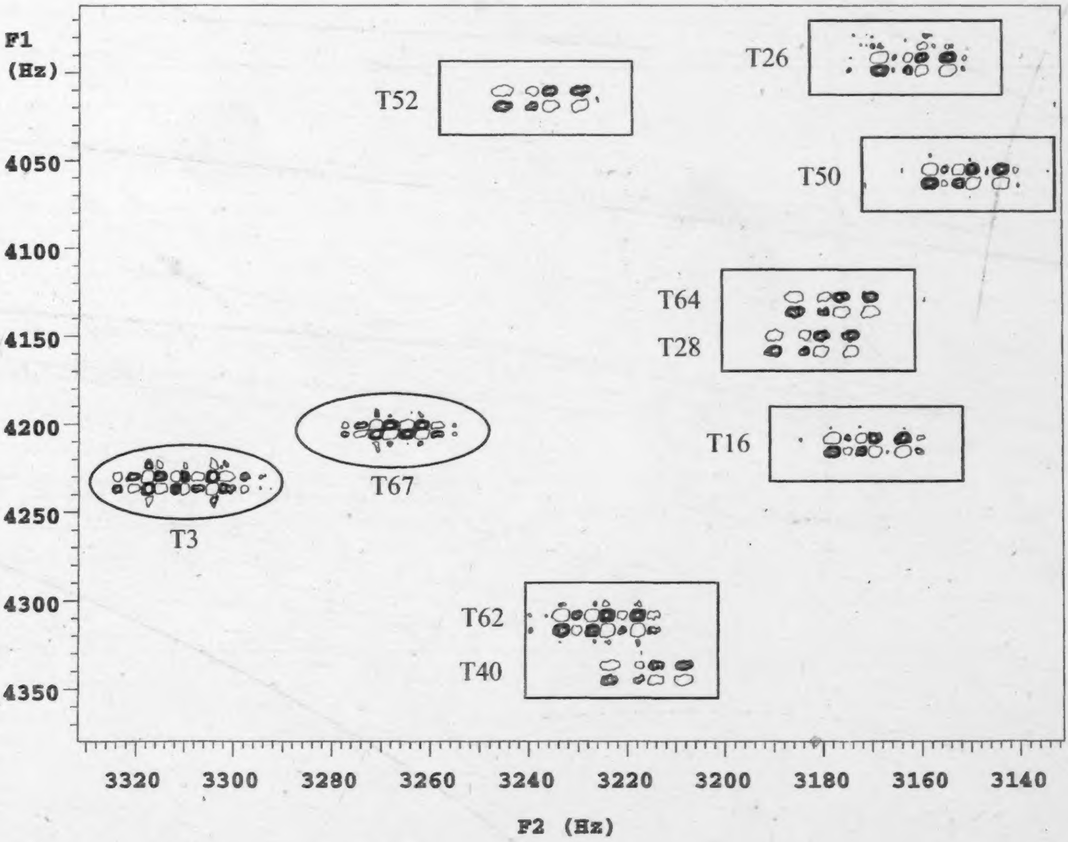


Figure 3.3: 2D DQF-COSY ^{1}H NMR spectrum at 800 MHz of TmAFP at 30 °C in 100% D_2O. Part of the Hα-Hβ region containing the threonine cross-peaks is displayed. The boxed threonines are those on the ice-binding face, while the circled ones are not.

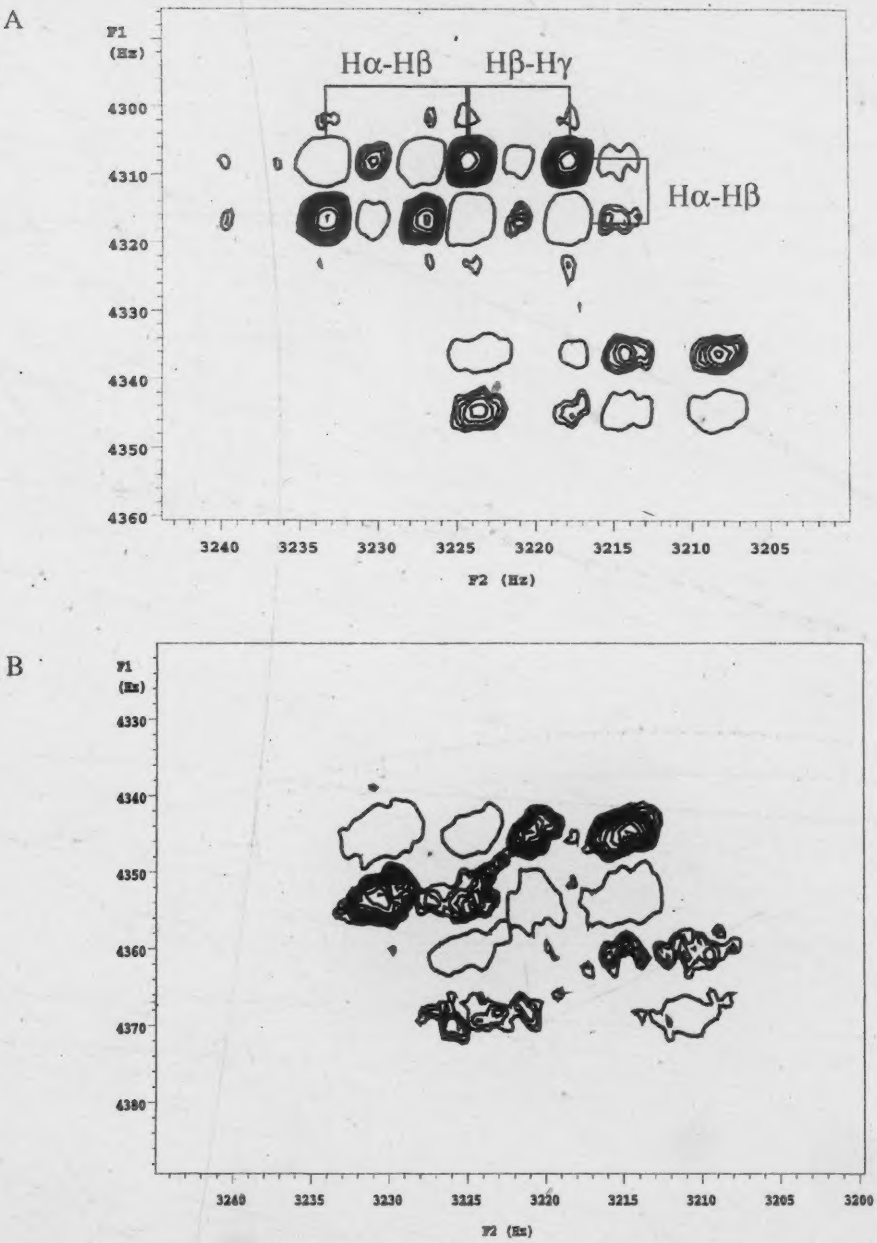


Figure 3.4: Expansions of the 2D DQF-COSY ^{1}H NMR spectra at 800 MHz of TmAFP at (A) 30 °C and (B) 15 °C. The region containing the Hα-Hβ cross-peak patterns of T40 (bottom) and T62 (top) are displayed to show the fine splitting of these cross-peaks. To distinguish between positive and negative peaks, the positive peaks are displayed as contoured peaks and the negative peaks are displayed as open circles.

A crosspeak between a pair of protons is due only to the coupling between them (active coupling); however, each proton in the pair may also have couplings to other nuclei (passive couplings). In a phase sensitive COSY experiment, these are indicated by the active couplings giving rise to splittings that are alternating in phase, while the passive couplings split the line without altering the phase. An example of a well-resolved crosspeak (T62) and its fine structure is shown in detail in Figure 3.4A. Despite the high digital resolution of these spectra, accurate values for the coupling constants are difficult to obtain directly, especially at the lower temperatures (Figure 3.4B), where the slower molecular tumbling results in increased line-broadening. As linewidths increase, the separation between the antiphase components, that is the active coupling being measured, becomes increasingly different from the actual splitting (*29*). To minimize this effect, the method of Kim and Prestegard (*30*) was used to measure peak-to-peak separations of the extrema in absorptive and dispersive plots of rows through the crosspeaks of interest. From these, an accurate calculation of scalar couplings can be made by analytical solution of the equations for Lorentzian lines. While the increased linewidths observed at lower temperatures are an issue, this method is highly accurate when the linewidths are less than J, becoming less accurate for linewidths greater than J (*30*). Owing to the extremely stable behaviour of TmAFP in solution and at lower temperatures, and the high digital resolution of the spectra, the extracted linewidths range from 2.1 - 4.2 Hz at 30 °C, and 5.2 - 8.8 Hz at 5 °C. These values for linewidths are similar to those obtained by spectral simulation of one-dimensional spectra. The values of Δv are less than J for those threonine residues on the ice binding face and on the order of J for the threonines not on the ice binding face. This implies that the observed ${}^3J_{\alpha\beta}$ coupling constants are accurate for residues on the ice-binding face and that the smaller values of J (that is the residues not on the ice-binding face) are affected by line broadening to a greater degree. This leads to an approximately 0.5 Hz overestimation of the ${}^3J_{\alpha\beta}$ measurements for the smallest values. A series of simulated crosspeak spectra generated using the program Mathematica (*31*), with the ${}^3J_{\beta\gamma}$ passive coupling set to 7 Hz (to indicate free rotation of the threonine methyl group) and the ${}^3J_{\alpha\beta}$ active coupling allowed to vary from 3 to 11 Hz, showed that the calculated splitting patterns matched those in the experimental spectrum.

The chemical shifts obtained at the various temperatures did not change significantly, however the coupling constants displayed a marked difference depending on their location in the protein and increased significantly as the temperature was lowered. The observed coupling constants and their temperature dependence are summarized by the graph in Figure 3.5. The threonine residues can clearly be separated into two distinct populations on the basis of these measured coupling constants. These are displayed as the blue (on the ice-binding face) and red (not on the ice-binding face) populations on the graph.

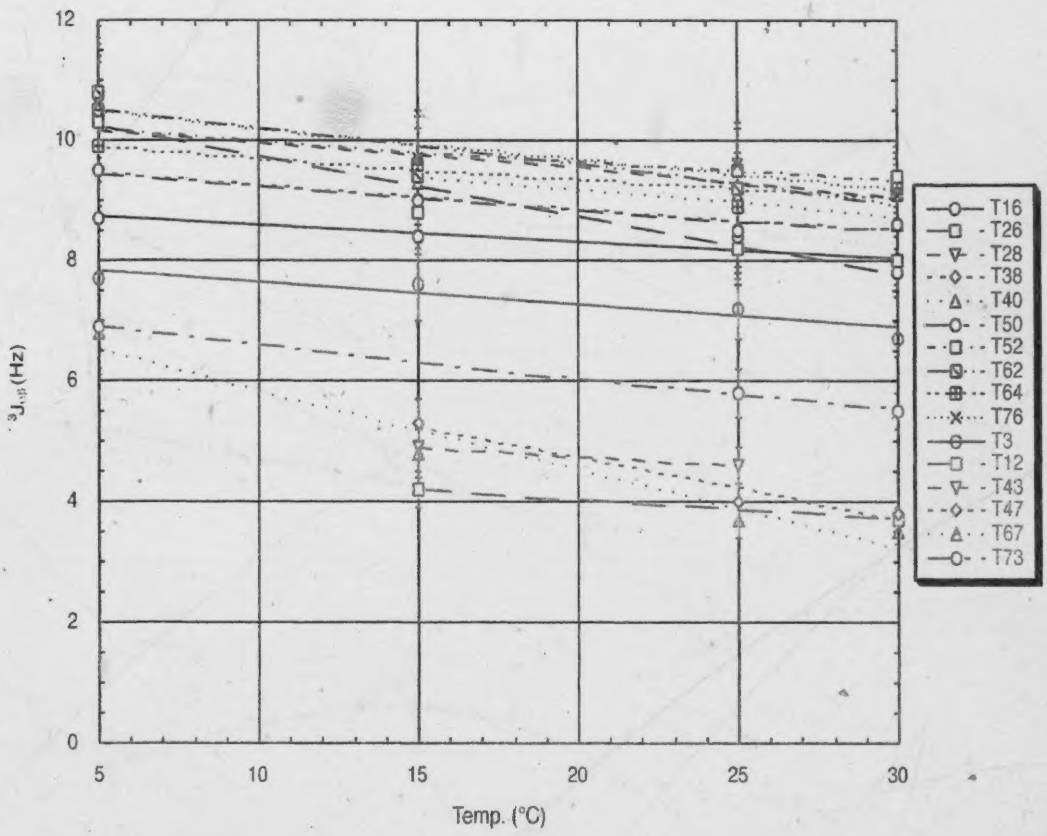


Figure 3.5: Plot of $^3J_{\alpha\beta}$ of each threonine residue as a function of temperature. The two observed populations correspond to those threonines on the ice-binding surface, in blue, and the threonines not on the ice-binding surface, in red.

The measured coupling constants are used to calculate the occupancy of a specific rotamer conformation on the basis of the expected coupling constants. In Figure 3.6, the expected coupling constants are displayed for the three staggered rotamer conformations of threonine. A side chain which does not have a fixed conformation, but instead undergoes rotation that is fast on the NMR time scale will have an average coupling constant of about 6.6 Hz.

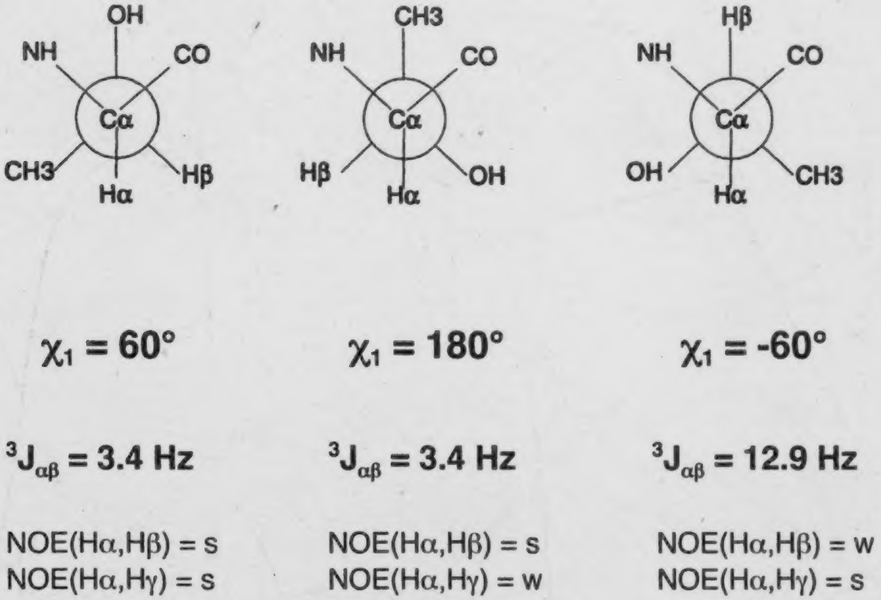

OH
NH
CO
Cα
CH3
Hα
Hβ
CH3
NH
CO
Cα
Hβ
Hα
OH
Hβ
NH
CO
Cα
OH
Hα
CH3
$\chi_1 = 60°$
$\chi_1 = 180°$
$\chi_1 = -60°$
$^3J_{\alpha\beta} = 3.4$ Hz
$^3J_{\alpha\beta} = 3.4$ Hz
$^3J_{\alpha\beta} = 12.9$ Hz
NOE(Hα,Hβ) = s
NOE(Hα,Hγ) = s
NOE(Hα,Hβ) = s
NOE(Hα,Hγ) = w
NOE(Hα,Hβ) = w
NOE(Hα,Hγ) = s

Figure 3.6: Correlation of the $^3J_{\alpha\beta}$ coupling constants and the corresponding NOE intensities with the χ_1 side chain torsion angles for the three most populated threonine rotamer conformations. The threonine side chain is shown in the Newman projection with the Cα in front and the Cβ in the back.

The observed $^3J_{\alpha\beta}$ coupling constants of T3 (6.7 Hz) and T73 (5.5 Hz) at 30 °C are very close to this expected value for unrestricted rotation and suggest either an equal population of all three rotamers or averaging between a gauche and the trans rotamer. Using T52, from one of the central loops of the ice-binding face as a representative ice-binding threonine, a coupling constant of 9.4 Hz is observed at 30 °C. This value increases to an observed coupling constant of 10.8 Hz at 5 °C. In the following equation,

P_1, P_2, and $1-P_1-P_2$ describe the fractional occupancy of the side chain in the +60°, 180°, and -60° conformations, respectively.

$$9.4\text{Hz} = P_1 \times 3.4\text{Hz} + P_2 \times 3.4\text{Hz} + \left(1 - P_1 - P_2\right) \times 12.9\text{Hz} \qquad (1)$$

Solving this equation shows that the side chain of T52 populates the χ_1=-60° conformation 63% of the time at 30 °C. At 5 °C, this increases to 73% population of the χ_1=-60° rotamer. These calculations were performed for all of the observed coupling constants at all four temperatures and these results are displayed in Table 3.1. The NOE analysis, as described in the Materials and Methods section, corroborates this observation of restricted rotation for the threonines on the ice-binding face.

Table 3.1: Percent population of the χ_1 = -60° rotamer conformation

Residue	30°C	25°C	15°C	5°C
Type I AFP[a]: T2				39
T13				54
T24				54
T35				34
TmAFP: T3[b]	35	40	44	45
T12[c]	3	-	9	-
T43[c]	-	13	16	-
T47[c]	5	7	20	-
T67[c]	1	4	15	36
T73[b]	22	25	-	37
T16	46	53	53	56
T26	48	50	57	75
T28	62	60	63	74
T38	55	54	59	64
T40	55	60	62	-
T50	55	54	59	64
T52	63	64	65	73
T62	60	61	63	78
T64	61	58	65	69
T76	61	65	67	76

[a] Values for Type I AFP are from Gronwald et al. 1996. [b] Values are consistent with free rotation of the sidechain. [c] Values are consistent with the χ_1 = +60° rotamer conformation.

In constrast, for T47, only 5 residues prior to T52 but not an ice-binding residue, a coupling constant of only 3.8 Hz is observed at 30 °C. This corresponds very nearly to complete occupancy of either the χ_1=180° or +60° rotamer, with only 5% population of the χ_1=-60° rotamer. The NOE analysis in this case supports the χ_1=+60° rotamer as the NOE ratio is close to 1. Upon lowering the temperature to 15 °C, the coupling constant does increase to 5.3 Hz, indicating that population of the χ_1=-60° rotamer is increasing at the expense of χ_1=+60°, suggesting the occurrence of rotameric averaging at the lower temperatures for the non ice-binding threonine residues. This cannot be distinguished on the basis of the NOE analysis since both a χ_1=+60° and a freely rotating side chain will have an NOE ratio of approximately 1. Nevertheless, it can be generally stated that the coupling constants for the ten threonines comprising the ice recognition surface populate the χ_1=-60° rotamer to a greater degree than the remaining threonines, which correspond to either an average of more than one conformation, or to preference for the χ_1=+60° rotamer.

3.3 Discussion

Although the ability of AFPs to bind ice and inhibit its growth is well-known, it is not well understood. Study of the molecular mechanism by which these proteins adsorb to ice are hindered by two main factors: first, the wide structural variation of AFPs for which structures have been solved, and second, the unique characteristics of the AFP-ice interaction which make it difficult to examine directly. The structural diversity of AFPs and the inability to deduce any sequence similarity between different AFP types means that putative ice-binding faces have been proposed based on mutation studies. The only features these varied sites appear to share is the relative flatness of the interaction surface of the protein, and in the case of Type I and the insect AFPs, the evenly spaced ranks of threonine residues that match the ice lattice. As well, a significant proportion of the surface area of the protein is involved in the binding interface, and the ice-binding sites generally appear to contain more hydrophobic and less polar amino acid side chains than the rest of the protein. Mutation studies of Type I and Type III fish AFPs and the β-helical insect AFPs have been very revealing, however. Mutations that introduce large side chains on the ice-binding surface sterically block the AFP-ice interaction, but are well-tolerated elsewhere on the protein (*23*). Mutations that decreased the size of the side chain are also deleterious, however mutations that are isosteric, that is those that do not change the shape or size of the side chain are generally neutral (*20*). This suggests that loss of shape complementarity and with it, the hydrophobic and van der Waals contacts, is an important factor contributing to the ability of these proteins to bind to ice. Mechanistically, for the adsorption-inhibition model of crystal growth to explain the observed properties of AFPs, the proteins must bind irreversibly. Therefore the antifreeze must fit almost perfectly to the ice in order to prevent water molecules from entering the interface (*32*). This can explain why any mutation that disrupts the "snugness" of the binding interface, whether larger or smaller, will have negative effects on ice-binding and thermal hysteresis.

In order to elucidate the molecular mechanism of ice binding, it is necessary to characterize the conformations of the critical side chains. The NMR analyses of TmAFP indicate that the ice-binding threonine residues have a clear preference for the χ_1=-60° rotamer conformation at all temperatures studied and that this preference becomes more pronounced at the lower temperatures, where the AFP is closer to physiologically relevant conditions. The threonine residues that are not on the ice-binding surface do not display this preferred conformation. This indicates that the TmAFP adopts a unique preformed ice-binding structure in solution prior to recognition of the ice surface. The rigidity of the side chains in the binding site reduces the entropic barrier to binding, as the side chains do not have to reposition themselves prior to ice binding. The experimental observation that the key binding side chains of TmAFP are positioned in the rotamer conformation adopted in the interface supports a hypothesis based on molecular dynamics simulations of proteins alone and in complex. In this study, it was proposed that this allows biological recognition to proceed on a feasible time scale and yields an intermolecular affinity that reduces the entropic penalty of binding (*33*). The flexibility and free rotation of the threonine residues not on the ice-binding face may help prevent the structuring of water molecules on the other side of the protein and prevent its engulfment into the ice crystal.

The β-helical structure of TmAFP provides the ideal scaffold to form this rigid ice binding conformation. The repetitiveness of the primary amino acid sequence is reflected in the three-dimensional loop structure allowing a rigid array of threonine residues to perfectly match the ice crystal lattice. Notably, the χ_1=-60° rotamer conformation is a commonly preferred rotamer in β-sheet secondary structure due to favorable hydrogen bonding effects with the backbone (*34*).

The inability to directly observe the molecular interaction of AFP and ice requires the use of various methods to indirectly probe the surface interaction. Computational docking studies and molecular dynamics simulations are a popular way to examine this interface, however the potential flexibility of protein surface side chains has posed a challenge for correct modeling. With this study, we have shown that the threonine side chains in the ice-binding face are not particularly flexible, and furthermore, we have determined the rotamer conformation most preferred by these side chains. This knowledge will contribute to the ability to perform more accurate molecular dynamics simulations and docking studies. Finally, the indirect probe we have employed, specifically solution state NMR experiments to study protein dynamics, has proven very useful in characterizing the rigid nature of the AFP-ice interface and the behaviour of the side chains in particular. Clearly the intimate surface-to-surface complementarity between TmAFP and crystalline ice is a key factor in binding.

3.4 Materials and Methods

3.4.1 NMR spectroscopy

An unlabeled sample of TmAFP isoform 2-14 previously prepared was used for all experiments described in this paper. This sample was prepared by dissolving the lyophilized protein in 100% D_2O, with 0.1 mM DSS added for internal referencing. The final protein concentration was approximately 2 mM and the pH was adjusted to 5.6 with microliter aliquots of 100 mM NaOD or DCl as required. For the freezing experiment, a medium-walled tube (524-PP-8, Wilmad, Buena, NJ) was used to prevent tube breakage.

The freezing experiment was collected on a Varian Unity 300 MHz spectrometer using an indirect detection probe. The ^{1}H 1D NMR spectra were collected over a temperature range of 30 °C to -20 °C in 1° decrements. After each temperature change, the sample was allowed to equilibrate for 30 minutes. For each temperature point, 16 000 complex data points were acquired with 256 transients using a spectral width of 4000 Hz. The 90° pulse width was calibrated to 7.5 μs.

The 2D DQF-COSY (*35*) and NOESY (*36*) spectra were acquired at 30 °C on a Varian INOVA 800 MHz spectrometer equipped with a 5-mm triple resonance probe and x, y, and z-axis pulsed field gradients. For the DQF-COSY experiment a spectral width of 7000 Hz was used in both dimensions. The acquired data consisted of 8192 F_2 × 1024 F_1 complex data points. The data were zero-filled to give a spectrum after transformation that contained 16 384 × 4096 data points. The 2D NOESY spectrum was acquired with a spectral width of 10 000 Hz in both dimensions and a mixing time of 100 ms. The data consisted of 8192 F_2 × 512 F_1 complex data points and was zero-filled to give a spectrum that contained 16 384 × 4096 data points after transformation.

All 2D DQF-COSY spectra collected for measurement of $^3J_{\alpha\beta}$ coupling constants were also measured on the Varian INOVA 800 MHz spectrometer. Spectra were acquired at 5, 15, 25, and 30 °C with a spectral width of 2000 Hz in F_2 and 2400 Hz in F_1. Temperature calibration of the spectrometer indicates the reported temperatures are within 0.5 °C from 5 to 20 °C and within 0.1 °C from 25 to 30 °C. The data consisted of 8192 F_2 × 1024 F_1 complex data points. The spectra were processed with 16 384 × 4096 data points using an unshifted sine bell window function in combination with line broadening in both dimensions. All NMR data processing, including the integration of the NOEs, was performed using the Varian VNMR 6.1B processing software on a Sun Blade 100 workstation.

3.4.2 Determination of $^3J_{\alpha\beta}$ coupling constants

The $^3J_{\alpha\beta}$ coupling constants of the threonine residues were determined at four temperatures from the peak separation of the DQF-COSY Hα - Hβ cross peaks. Even with the high digital resolution of the spectra, it is difficult to obtain accurate values for coupling constants directly from the observed splitting as the separation between antiphase components becomes increasingly different from the actual splitting with increasing linewidths (*29*). Slower molecular tumbling at decreased temperatures causes this effect to become more pronounced at the lower temperatures measured. For this reason we used the method of Kim and Prestegard (*30*), which allows accurate calculation of scalar couplings by the measurement of peak-to-peak separations of the extrema in absorptive and dispersive plots of rows through the cross peaks of interest. These peak-to-peak separations were measured along the higher resolution F_2 axis in the opposite trace orientation from that displayed in Figure 3.4A to reduce partial overlap in the multiplet.

3.4.3 Determination of χ_1 side chain torsion angles

The χ_1 side chain torsion angles are obtained by the analysis of the pattern of ${}^3J_{\alpha\beta}$ coupling constants and the relative intensities of the intraresidue NOEs involving the Hα and two Hβ protons (*37-39*). Figure 3.6 displays the necessary information for determining these angles for the simpler case of threonine residues, which have only one Hβ proton. From the observed coupling constant, it is possible to calculate the occupancy of a specific rotamer conformation on the basis of the expected coupling constants for the different staggered rotamers.

$$J_{\mathrm{obs}} = \sum_i P_i J_i \tag{2}$$

where P_i is the fractional occupancy of a conformation corresponding to the coupling constant J_i. For a side chain which is not fixed into one conformation, an average coupling constant will be obtained if the rotation is fast on the NMR time scale. In the case of a threonine side chain with no rotamer preference, a ${}^3J_{\alpha\beta}$ of 6.6 Hz would be obtained. In addition to the information about preferred rotameric states obtained from coupling constant analyses, the ratio of NOE intensities of the α-, β-, and γ-protons provides further evidence. The NOE ratio is defined by:

$$\text{ratio} = \frac{\left(\mathrm{NOE}_{\mathrm{H}\alpha\text{-}\mathrm{H}\beta}\right)}{\frac{1}{3}\left(\mathrm{NOE}_{\mathrm{H}\alpha\text{-}\mathrm{H}\gamma}\right)} \tag{3}$$

The Hα-Hγ NOE is divided by 3 to account for the three threonine methyl protons as compared to the one β-proton. As outlined in Figure 3.6, a NOE ratio significantly less than 1 corresponds to a χ_1 of -60°, while a NOE ratio significantly greater than 1 corresponds to a χ_1 of 180°. A NOE ratio of approximately 1 corresponds to either a χ_1 of +60° or to a freely rotating side chain.

3.5 Serine and Valine Mutation Studies*

3.5.1 Rationale

As mentioned briefly in the Introduction to this chapter, one of the initial suggestions for the ice-binding mechanism was that the polar side chains of an AFP hydrogen bond to the ice surface. The model was based on the crystal structure of the well-studied Type I AFP from winter flounder. This α-helical AFP is composed of 37 residues with a tandemly repeated 11 amino acid consensus sequence TX_2N/DX_7, where X is generally alanine (*14*).

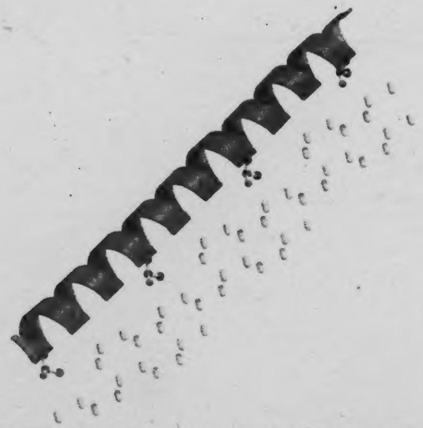


Figure 3.7: Type I AFP crystal structure (*13*; PDB 1WFB, chain B) docked to the pyramidal plane of ice. The side chains of the four threonine residues are displayed in ball-and-stick representation. (Figure adapted from *40*).

---

* This work was presented in a poster at the PENCE Annual General Meeting, Vancouver, British Columbia, May 22-23, 2003. Daley, M.E., Marshall, C.B., Davies, P.L., and Sykes, B.D. Side chain conformational flexibility of *Tenebrio molitor* antifreeze protein and its role in ice surface recognition.

The lattice distance matching model proposed that hydrogen bonding would occur between the four threonine hydroxyl groups and the ice, based on the observation that the 16.5 Å spacing of the *i*, *i*+11 threonine residues matched the 16.7 Å distance of water molecules on the pyramidal plane (*11*). Further experimentation has led to a diminished importance for the hydrogen bonding interaction. Specifically, mutation of the two central threonine residues to serine caused a severe loss of activity (~10% of wild type activity), despite preserving the hydroxyl group and therefore the hydrogen bonding activity of the side chain. In comparison, mutation of the two central threonine residues to valine resulted in only a minor loss of activity (~85% of wild type activity) by preserving the methyl group and overall shape of the side chain (*15*).

With the knowledge that the TmAFP ice binding residues are also regularly arrayed threonine residues, we decided to attempt a similar study with single serine and valine mutations on the ice-binding face. We then applied the same method to analyze the rotamer conformations of the remaining threonine side chains using the $^3J_{\alpha\beta}$ coupling constants as previously described in this chapter.

3.5.2 Experimental procedures

Unlabeled wild-type TmAFP isoform 4-9, and single mutations T16S, T16V, T38S, T38V, T40S, and T40V, were produced in the manner previously described (*23*). For a justification of using the 4-9 isoform in place of 2-14, refer to Appendix C. Thermal hysteresis measurements on the mutant proteins have been performed, and all the serine and valine mutations cause some degree of loss of activity relative to wild type TmAFP (C.B. Marshall and P.L. Davies, personal communication). The $^3J_{\alpha\beta}$ ^1H-^1H scalar coupling constants were measured using the same procedure as the wild type 2-14 coupling constants described in section 3.4.

3.5.3 Results

The experiments were collected at 600 MHz with high digital resolution, however the quality of the DQF-COSY spectra varied because the protein concentrations were not uniform. Therefore, no coupling constant measurements could be made for T16V and some samples are lacking measurements at various positions. To obtain data comparable to the 2-14 isoform measurements already obtained, we also collected higher resolution DQF-COSY experiments of the WT 4-9 isoform and the T40 mutant proteins at 800MHz. These measurements are displayed in Table 3.2 (for the 600 MHz data) and Table 3.3 (for the 800 MHz data).

Table 3.2: Threonine $^3J_{\alpha\beta}$ coupling constants (Hz) measured at 600 MHz and 25°C

Residue	WT	T16S	T38S	T38V	T40S	T40V
T3	6.9	6.8	6.9	7.0	7.1	6.9
T16	7.7	S	7.8	7.3	7.9	7.9
T26	7.4	7.2	6.8	-	-	6.9
T28	9.1	9.2	9.5	9.0	8.9	9.4
T38	8.1	8.7	S	V	7.9	7.7
T40	8.6	8.3	8.1	-	S	V
T52	9.3	9.3	9.3	8.3	8.1	8.8
T62	9.0	9.2	8.7	8.6	9.0	9.0
T64	8.6	8.3	8.8	-	8.9	9.4
T67	5.9	-	6.0	-	-	-
T76	8.8	9.1	8.9	9.0	9.0	8.9

Table 3.3: $^3J_{\alpha\beta}$ coupling constants (Hz) of T40 mutations measured at 800 MHz and 25°C

Residue	WT	T40S	T40V
T3	6.8	7.2	7.4
T16	8.5	7.5	8.1
T26	7.9	6.8	6.8
T28	8.8	9.1	8.7
T38	8.2	7.8	8.1
T40	9.0	S	V
T50	8.2	7.8	8.1
T52	9.4	7.9	9.5
T62	8.8	8.8	9.0
T64	8.8	8.9	9.1
T67	5.2	3.7	3.4
T76	9.5	9.1	9.2

A plot of the data in Table 3.3 is provided in Figure 3.8, with error bars included to show the estimated 0.5 Hz error in the measurements. The boxed values in Tables 3.2 and 3.3 correspond to those ice binding residues for which a nearby mutation to either serine or valine significantly decreased (ie. 1 Hz or greater) the measured $^3J_{\alpha\beta}$ coupling constant relative to the wild-type protein.

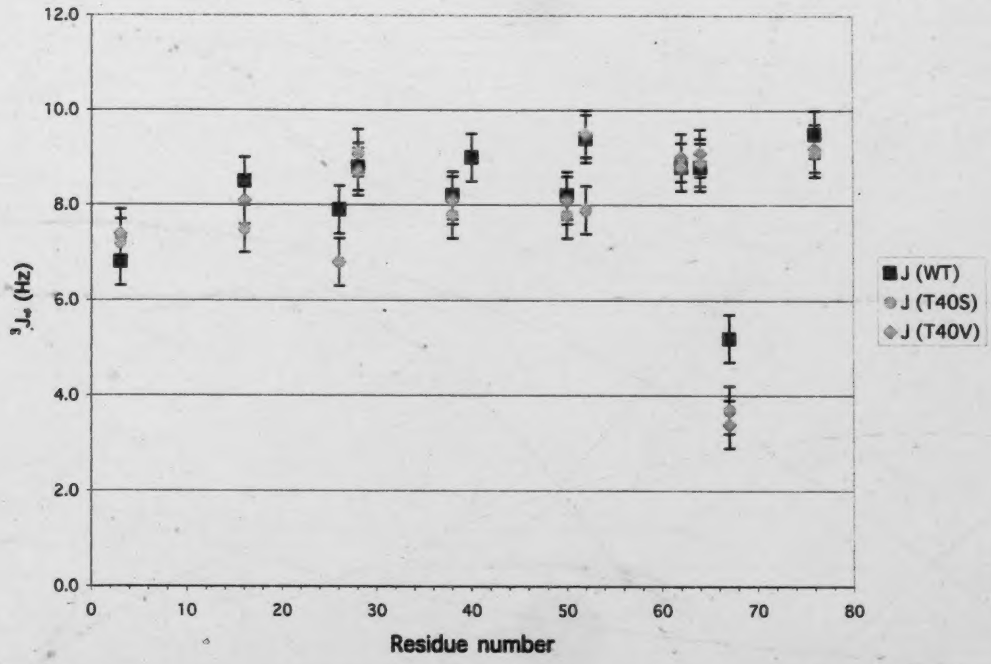


Figure 3.8: Plot of $^3J_{\alpha\beta}$ in Hz for the threonine residues of the wild-type, T40S and T40V proteins. These were measured on a Varian 800 MHz spectrometer at 25°C.

3.5.4 Discussion

First, we will examine the measurements for the 600 MHz data. Mutations in T38 and T40 represent two threonines in the same, central loop of TmAFP, but in a different rank of threonine residues. For the view of the ice binding face in Figure 3.1B, T38 is in the right hand rank containing four threonines, while T40 is in the left hand rank, which is longer and contains six threonine residues. T52, in the loop below these mutations, is affected by certain mutations as shown by the decrease in the measured coupling constant. The effects of the mutations appear to be specific to the position of the threonine. For T38, mutation to valine, but not to serine, decreases the $^3J_{\alpha\beta}$ coupling constant of T52, however for T40, only mutation to serine has the equivalent negative effect.

Upon examination of the T40 mutants in greater detail using the 800 MHz data, differences were observed in the pattern of how surrounding residues are affected depending on the type of mutation made. For T40S, the coupling constants of three residues were significantly decreased: T16, T26, and T52. For T40V, only T26 was observed to have a significantly lower $^3J_{\alpha\beta}$ measurement.

A decrease in the measured coupling constant is an indication that the preference of that side chain for the χ_1=-60° rotamer population is decreased. These results are complicated and no simple pattern emerges, but we suggest that mutation of ice-binding threonine residues to serine or valine alters the positioning of neighbouring threonine residues. These effects appear to be slightly different depending on the nature of the mutation made, as well as the position of the mutated threonine.

In general, this can be viewed as a destabilization of the rigid array of threonines on the ice binding face of TmAFP. This observation lends further experimental support to the hypothesis that intimate surface-to-surface complementarity between the AFP and crystalline ice is a key factor in ice binding and inhibition of ice crystal growth.

3.6 References

1. Fletcher, G.L., Hew, C.L., and Davies, P.L. (2001) *Annu. Rev. Physiol.* **63:** 359-390.

2. Duman, J.G. (2001) *Annu. Rev. Physiol.* **63:** 327-357.

3. Worrall, D., Elias, L., Ashford, D., Smallwood, M., Sidebottom, C., Lillford, P., Telford, J., Holt, C., and Bowles, D. (1998) *Science* **282:** 115-117.

4. Sidebottom, C., Buckley, S., Pudney, P., Twigg, S., Jarman, C., Holt, C., Telford, J., McArthur, A., Worrall, D., Hubbard, R., and Lillford, P. (2000) *Nature* **406:** 256.

5. Jia, Z. and Davies, P.L. (2002) *TIBS* **27:** 101-106.

6. Graham, L.A., Liou, Y.-C., Walker, V.K., and Davies, P.L. (1997) *Nature* **388:** 727-728.

7. Davies, P.L. and Sykes, B.D. (1997) *Curr. Opin. Struct. Biol.* **7:** 828-834.

8. Raymond, J.A. and DeVries, A.L. (1977) *Proc. Natl. Acad. Sci. USA* **74:** 2589-2593.

9. Houston, Jr., M.E., Chao, H., Hodges, R.S., Sykes, B.D., Kay, C.M., Sönnichsen, F.D., Loewen, M.C., and Davies, P.L. (1998) *J. Biol. Chem.* **273:** 11714-11718.

10. Graether, S.P., Kuiper, M.J., Gagné, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. (2000) *Nature* **406:** 325-328.

11. Knight, C.A., Cheng, C.C., and DeVries, A.L. (1991) *Biophys. J.* **59:** 409-418.

12. Wen, D. and Laursen, R.A. (1992) *Biophys. J.* **63:** 1659-1662.

13. Sicheri, F. and Yang, D.S.C. (1995) *Nature* **375:** 427-431.

14. Davies, P.L. and Hew, C.L. (1990) *FASEB J.* **4:** 2460-2468.

15. Chao, H., Houston, Jr., M.E., Hodges, R.S., Kay, C.M., Sykes, B.D., Loewen, M.C., Davies, P.L., and Sönnichsen, F.D. (1997) *Biochemistry* **36:** 14652-14660.

16. Haymet, A.D.J., Ward, L.G., Harding, M.M., and Knight, C.A. (1998) *FEBS Lett.* **430:** 301-306.

17. Haymet, A.D.J., Ward, L.G., and Harding, M.M. (1999) *J. Am. Chem. Soc.* **121:** 941-948.

18. Zhang, W. and Laursen, R.A. (1998) *J. Biol. Chem.* **273:** 34806-34812.

19. Baardsnes, J., Kondejewski, L.H., Hodges, R.S., Chao, H., Kay, C., and Davies, P.L. (1999) *FEBS Lett.* **463:** 87-91.

20. Baardsnes, J. and Davies, P.L. (2002) *Biochim. Biophys. Acta* **1601:** 49-54.

21. Gronwald, W., Chao, H., Reddy, D.V., Davies, P.L., Sykes, B.D., and Sönnichsen, F.D. (1996) *Biochemistry* **35:** 16698-16704.

22. Liou, Y.-C., Tocilj, A., Davies, P.L., and Jia, Z. (2000) *Nature* **406:** 322-324.

23. Marshall, C.B., Daley, M.E., Graham, L.A., Sykes, B.D., and Davies, P.L. (2002) *FEBS Lett.* **529:** 261-267.

24. Daley, M.E., Spyracopoulos, L., Jia, Z., Davies, P.L., and Sykes, B.D. (2002) *Biochemistry* **41:** 5515-5525.

25. Kraulis, P.J. (1991) *J. Appl. Crystallogr.* **24:** 946-950.

26. Merritt, E.A. and Bacon, D.J. (1997) *Methods Enzymol.* **277:** 505-524.

27. Liou, Y.-C., Daley, M.E., Graham, L.A., Kay, C.M., Walker, V.K., Sykes, B.D., and Davies, P.L. (2000) *Protein Exp. Purif.* **19:** 148-157.

28. Graether, S.P., Slupsky, C.M., Davies, P.L., and Sykes, B.D. (2001) *Biophys. J.* **81:** 1677-1683.

29. Wüthrich, K. (1986) *NMR of proteins and nucleic acids.* John Wiley & Sons, New York.

30. Kim, Y. and Prestegard, J.H. (1989) *J. Magn. Reson.* **84:** 9-13.

31. Wolfram, S. (1996) *Mathematica version 3.0.* Wolfram Research Inc.

32. Knight, C.A. and Wierzbicki, A. (2001) *Cryst. Growth Des.* **1:** 439-446.

33. Kimura, S.R., Brower, R.C., Vajda, S., and Camacho, C.J. (2001) *Biophys. J.* **80:** 635-642.

34. Dunbrack, Jr., R.L. and Karplus, M. (1994) *Nat. Struct. Biol.* **1:** 334-340.

35. Rance, M., Sørensen, O.W., Bodenhausen, G., Wagner, G., Ernst, R.R., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **117:** 479-485.

36. Jeener, J., Meier, B.H., Bachmann, P., and Ernst, R.R. (1979) *J. Chem. Phys.* **71:** 4546-4553.

37. Karplus, M. (1959) *J. Chem. Phys.* **30:** 11-15.

38. Karplus, M. (1963) *J. Am. Chem. Soc.* **85:** 2870-2871.

39. Clore, G.M. and Gronenborn, A.M. (1989) *Crit. Rev. Biochem. Mol. Biol.* **24:** 479-557.

40. Graether, S.P. (1999) *PhD Thesis.* Queen's University, Kingston, Ontario.

CHAPTER 4: Characterization of Threonine Side Chain Dynamics Using Natural Abundance ^{13}C NMR Spectroscopy[*]

4.1 Introduction

The study of protein dynamics is of fundamental importance towards a complete understanding of their biological function. Nuclear magnetic resonance (NMR) spectroscopy is widely used to examine the internal motions of proteins over a range of time-scales, including large amplitude slow motions (seconds), intermediate and exchange motions (milli- to microseconds), and fast bond vector fluctuations (nano- to picoseconds) all of which are significant in various biological contexts (*1*). The heteronuclear NMR relaxation techniques most often employed are sensitive to these subnanosecond internal motions. With increasing evidence that picosecond protein dynamics may be relevant to the retention of entropy in protein-ligand binding (*2*), measurements of this type will further contribute to a full description of the relationship between protein structure and function.

The widespread use of ^{15}N NMR relaxation measurements has provided a great deal of information on protein backbone dynamics, however a complete description requires information about side chain motions as well. For this, the use of ^{13}C spectroscopy is necessary. Side chain dynamics have been examined by various methods, both on proteins that are uniformly ^{13}C/^{15}N labeled (*3, 4*) or those that are partially labeled by either random fractional ^{13}C labeling (*5*) or the incorporation of ^{13}C into alternating carbon positions (*6*), both of which avoid complicated data interpretation arising from the large ^{13}C-^{13}C scalar and dipolar couplings. The incorporation of deuterium as a spin relaxation probe at various side chain positions (*7, 8*) has also proven extremely valuable.

---

[*] A version of this chapter has been submitted for publication. Daley, M.E. and Sykes, B.D. Characterization of threonine side chain dynamics in an antifreeze protein using natural abundance ^{13}C NMR spectroscopy. *J. Biomol. NMR.*, submitted August 5, 2003.

In the absence of ^{13}C labeling, the use of natural abundance ^{13}C NMR spectroscopy has allowed the examination of protein motions despite sensitivity limitations (*9-12*). While many natural abundance ^{13}C NMR experiments have been successful, the techniques are best suited for methine carbon relaxation because the simpler system allows for greater ease of analysis. Methylene and methyl carbon relaxation by natural abundance methods is considerably more complex owing to cross-correlation effects (*13-15*). These techniques have not only enabled study of the motions of backbone Cα and side chain methine carbons, but have also permitted direct comparison to backbone ^{15}N measurements (*16*). In the case of the ^{15}N nucleus, relaxation occurs through both the dipolar interaction with its attached amide proton and its chemical shift anisotropy, while for natural abundance ^{13}C, relaxation of aliphatic carbons is dominated only by the dipole-dipole interaction with the directly attached proton. The advantage of focusing on these nearly ideal two-spin systems is that the heteronuclear relaxation is attributed to the reorientation of the HX bond vector with respect to the direction of the magnetic field and can be analyzed in terms of the local motions being superimposed on the overall protein tumbling. However, the choice of an appropriate model for the motion is often difficult and internal motion can be interpreted either using model-free approaches (*17-20*) or in terms of specific motional models including wobbling-in-a-cone (*21, 22*), site jump, or restricted diffusion models (*23*).

For the antifreeze protein from the beetle *Tenebrio molitor* (TmAFP), we have previously completed a full analysis of ^{15}N backbone relaxation parameters at both 30°C and 5°C (*24*). These studies showed that TmAFP is a well-folded and rigid β-helical protein with restricted internal mobility throughout the backbone. Owing to difficulties in the expression and purification of this highly disulfide-bonded protein (*25*), we have not been able to obtain ^{13}C labeled TmAFP. In order to study side chain motions in this protein, and in particular, the threonine side chains which comprise the ice-binding face (*26*), we have chosen to investigate the dynamics of the χ_1 dihedral angles connecting the β-CH of threonine to its corresponding α-CH. In our previous study of TmAFP side chains (*27*), we analyzed the internal dynamics of the threonine χ_1 angles using motionally averaged $^3J_{\alpha\beta}$ ^{1}H-^{1}H scalar coupling constants. Using the Karplus relations (*28, 29*) and the relative intensities of the intraresidue NOEs involving the Hα and Hβ protons (*30*), populations of the staggered rotamer conformations are calculated by fitting the $^3J_{\alpha\beta}$ coupling constant to the measured value. We observed that threonine residues on the ice-binding face preferentially maintain a χ_1=-60° rotamer conformation and concluded that a rigid preformed ice binding structure is adopted in solution prior to ice surface recognition. However, using averaged J coupling constants provides no information about the rates of transfer between rotameric states or of librational motions. In this paper, we examine the χ_1 dihedral angle dynamics using ^{13}C methine relaxation rates measured at natural abundance.

4.2 Experimental Procedures

4.2.1 NMR spectroscopy

Methine ^{13}C relaxation experiments were performed on a single previously prepared sample of TmAFP isoform 2-14 at natural ^{13}C abundance. The lyophilized protein was dissolved in 100% D_2O, with 0.1 mM DSS added for internal referencing. The final protein concentration was approximately 2 mM and the pH was adjusted to 5.6 with microliter aliquots of 100 mM NaOD or DCl as required. The NMR experiments were performed using Varian INOVA 500 MHz and Unity 600 MHz spectrometers equipped with 5 mm triple resonance probes and *z*-axis pulsed field gradients. ^{13}C-R_1, -R_2, and {^{1}H}-^{13}C NOE experiments were carried out at 125.7 and 150.8 MHz at both 30°C and 5°C using sensitivity-enhanced gradient HSQC pulse sequences adapted from those of Farrow et al. (*31*). The CH/CH_3 – edited HSQC spectrum of TmAFP is displayed in Figure 4.1 with assignments of CαH and CβH resonances labeled.

The spectral widths for ^{1}H and ^{13}C for the R_1 and NOE experiments at each temperature were 5500.0 and 5000.0 Hz at 500 MHz, and 6400.0 and 6000.0 Hz at 600 MHz, respectively. The ^{1}H carrier was set at the residual HDO resonance, and the ^{13}C carrier was set at 56.7 ppm. The number of real data points acquired for ^{1}H and ^{13}C, respectively, were 396 and 128 (R_1) or 396 and 48 (NOE) at 500 MHz, and 436 and 128

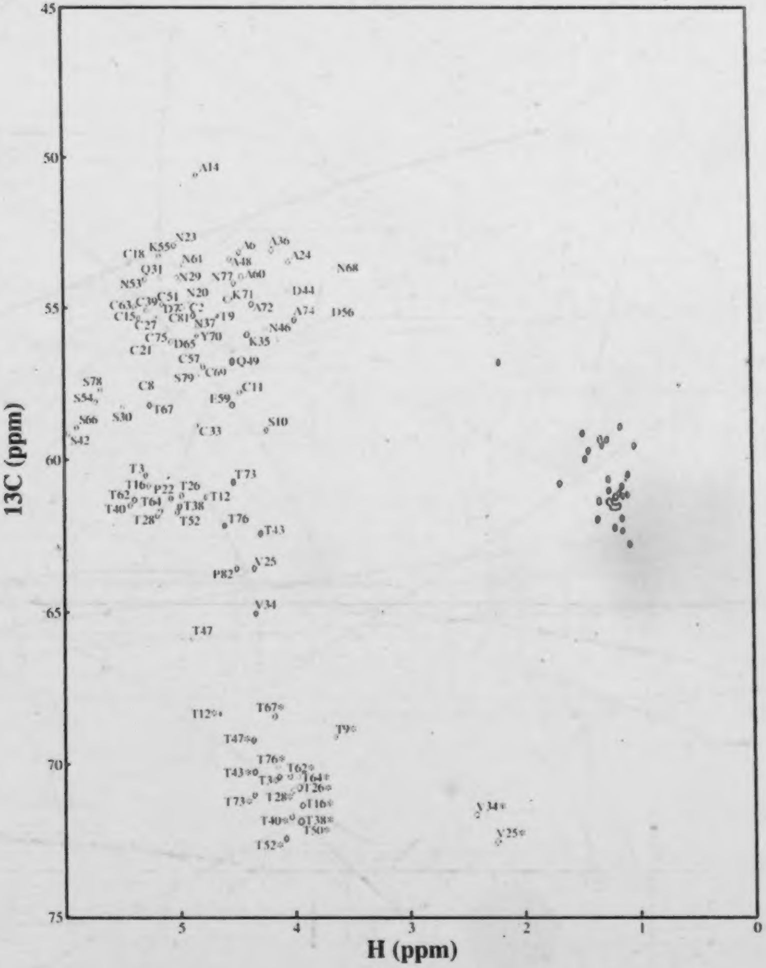


Figure 4.1: ^{1}H-^{13}C HSQC correlation spectrum of TmAFP methine and methyl resonances. The labels display the assignment of the CαH and CβH (indicated by asterisks) correlations.

(R_1) or 410 and 48 (NOE) at 600 MHz. A total of 48 (at 500 MHz) or 64 (at 600 MHz) transients were collected for ^{13}C-R_1. At 500 and 600 MHz, 256 transients were accumulated for the measurement of {^{1}H}-^{13}C NOE. For the measurement of ^{13}C-R_2 at 30°C, reduced spectral widths of 2837.081 Hz at 500 MHz and 3404.57 Hz at 600 MHz were used in the F_1 (^{13}C) dimension and the ^{13}C carrier was moved to 62.7 ppm to minimize off-resonance effects on the relaxation of the Cβ resonances.

For measurement of ^{13}C-R_1 relaxation rates, delays of 10.1, 50.5, 111.1, 181.8, 252.5, 353.5, 454.5, and 606.1 ms were used at both 500 and 600 MHz. At 5°C, delays of 10.1, 50.7, 111.5, 182.5, 253.5, 354.9, 456.3, and 608.4 ms were used. Measurement of ^{13}C-R_2 relaxation rates, at 30°C only, employed delays of 16.3, 32.6, 48.9, 65.2, 81.5, 97.8, 114.1, and 130.4 ms at both 500 and 600 MHz. In order to obtain equilibrium, a 2.0 s delay between repetitions of the pulse sequence for ^{13}C-R_1 measurements was employed, while the delay for obtaining equilibrium during the measurement of ^{13}C-R_2 was 6.5 s. $\{^1H\}$-^{13}C steady-state NOEs were measured from two HSQC spectra acquired with and without proton saturation prior to the first ^{13}C excitation pulse. Proton saturation at both 500 and 600 MHz was achieved using a train of 120° proton pulses with 5 ms pulse intervals for a total 3.5 s of saturation. Relaxation delays of 2 or 5 ms between repetitions of the pulse sequence were used in the experiments with and without proton saturation, respectively.

4.2.2 NMR data processing

All NMR data were processed with the NMRPipe software (*32*). Enhanced sensitivity data were processed using the ranceY.M macro. The F_1 (^{13}C) dimension was extended by 64 complex points using linear prediction before zero filling. The F_2 (^{1}H) dimension was multiplied by a 60°-shifted sine-bell function and the F_1 dimension was multiplied by a 60°-shifted squared sine-bell function before Fourier transformation. The F_1 and F_2 dimensions were baseline corrected by polynomial subtraction in the frequency domain. The NMRView program (*33*) was used for peak picking of all ^{13}C-HSQC spectra. The values of the peak intensities for the ^{13}C-R_1 and -R_2 measurements were fit to single-exponential, two-parameter decay curves using the available NMRView rate analysis scripts. Error in the ^{13}C-R_1 and -R_2 measurements was obtained from non-linear least-squares fits of the peak intensities to two-parameter exponential decays. Error in the $\{^1H\}$-^{13}C NOE values was calculated from baseplane noise values in ^{13}C-HSQC spectra acquired with and without proton saturation.

4.3 Theory

The choice of the theoretical model that best fits the measured relaxation parameters can be difficult, especially with a limited amount of data; in general the simplest model that adequately describes the measured parameters should be used. In this paper, we decided not to use the model-free approach for the analysis, but instead have chosen to interpret the relaxation data using the rotational restricted diffusion model (*34*), which allows coverage of a broad range of amplitudes and time scales around the χ_1 dihedral angle (*4*). This model allows for free internal diffusion over a restricted range subject to boundary conditions that limit the range of motion. Specifically, the ^{13}Cβ-^{1}Hβ bond vector moves around χ_1 with an amplitude in the range $-\gamma_0 \leq \gamma \leq \gamma_0$ (Figure 4.2).

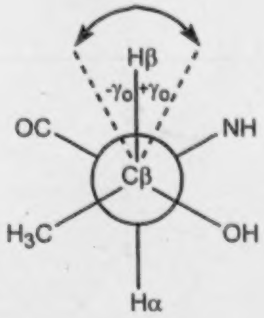


Figure 4.2: Restricted rotation of ±γ_0 around the Cα-Cβ bond. The threonine side chain is displayed as a Newman projection with the Cα in the back and the Cβ in the front. Rotation about the χ_1 angle is perpendicular to the plane of the page.

To examine the internal motion of the χ_1 dihedral angle fully, ^{13}C-R$_1$ and R$_2$ relaxation rates and the steady-state {^{1}H}-^{13}C NOE of the CH nuclei were measured. As mentioned briefly in the introduction, ^{13}C methine relaxation is caused by the dipolar interaction with its directly attached proton and the chemical shift anisotropy and cross-correlation between these two effects can safely be neglected (*35*). The expressions for the relaxation rate constants ^{13}C-R$_1$ and R$_2$ and the heteronuclear NOE are given by Abragam (*36*):

$$R_1 = \frac{\left(\frac{h}{2\pi}\right)^2 \gamma_H^2 \gamma_C^2}{4 r_{CH}{}^6}\left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)\right] \qquad (1)$$

$$R_2 = \frac{\left(\frac{h}{2\pi}\right)^2 \gamma_H^2 \gamma_C^2}{8 r_{CH}{}^6}\left[4J(0) + J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H) + 6J(\omega_H + \omega_C)\right] \qquad (2)$$

$$\text{NOE} = 1 + \frac{\gamma_H\left[6J(\omega_H + \omega_C) - J(\omega_H - \omega_C)\right]}{\gamma_C\left[J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)\right]} \qquad (3)$$

where γ_H, γ_C and ω_H, ω_C correspond to the gyromagnetic ratios and Larmor frequencies of hydrogen and carbon, respectively, *h* is Planck's constant and r_{CH} is the C-H bond length (1.10Å).

For the restricted diffusion motion about χ_1 in the range $\pm\gamma_0$, the correlation function is given by London and Avitabile (*34*):

$$C(t) = \frac{1}{4\pi} e^{-6Dt} \sum_{m=-2}^{2} \sum_{n=0}^{\infty} a_m \left(E_{m,n}(\gamma_0)\right)^2 e^{\frac{-D_1 n^2 \pi^2 t}{(4\gamma_0)}} \qquad (4)$$

where $D_1 = 1/\tau_i$.

The corresponding spectral density function is the Fourier transform:

$$J(\omega) = \sum_{m=-2}^{2} \sum_{n=0}^{\infty} a_m \left(E_{m,n}(\gamma_0)\right)^2 \frac{2\tau_{cn}}{1 + (\omega\tau_{cn})^2} \qquad (5)$$

where $E_{m,n}(\gamma_0)$ is defined by:

$$\begin{array}{ll} \dfrac{\sin(m\gamma_0)}{m\gamma_0} & \text{for n=0} \\[2ex] \dfrac{1}{\sqrt{2}}\left(\dfrac{\sin(m\gamma_0 - n\frac{\pi}{2})}{m\gamma_0 - n\frac{\pi}{2}} + (-1)^n \dfrac{\sin(m\gamma_0 + n\frac{\pi}{2})}{m\gamma_0 + n\frac{\pi}{2}}\right) & \text{for n}\neq 0 \end{array} \qquad (6)$$

and $$\tau_{cn} = \frac{4\gamma^2{}_0}{24D\gamma^2{}_0 + D_1 n^2 \pi^2} \qquad (7)$$

where $D = 1/6\tau_c$. The constants are given by $a_{\pm 2}=0.2961$, $a_{\pm 1}=0.1486$ and $a_0=0.1107$ (*4*).

London and Avitabile (*34*) conclude for this model, that T_2 values provide a more useful interpretive tool for internal motion in most cases. They also state that T_1 and the NOE are valid only for diffusion coefficients where the free internal rotation calculation is in reasonable agreement with the $\gamma_0=180°$ calculation and when their behaviour with γ_0 is nearly monotonic, which criteria are satisfied for small proteins as used herein. Engelke and Rüterjans (*4*) did not include T_2 values in the analysis of RNase T1 using this model, deriving qualitative lower limits for the amplitudes of motion from NOE values and including T_1 in the quantitative analysis to determine the upper limits for the internal correlation time. In the present paper, we will use both R_1 and R_2 relaxation rates as well as the NOE to estimate the amplitude of motion γ_0. The simulations were performed using the program Mathematica (*37*) using the equations described and the parameters detailed in the following sections.

4.4 Results

4.4.1 ^{13}C-R_1, -R_2, and NOE data

^{13}C methine relaxation parameters for TmAFP were studied at two temperatures, 30 and 5°C. At 30°C, the high resolution of the spectra allowed for 53-60 of 74 non-glycine Cα and 12-15 of 19 Cβ measurements at 600 MHz. At 500 MHz, 50-54 Cα and 10-13 Cβ resonances were used in the analysis. The values of ^{13}C-R_1, -R_2 and NOE at 30°C as a function of residue number are shown for the 500 MHz data in Figure 4.3 and the 600 MHz data in Figure 4.4. The ^{13}C-R_1, -R_2, and NOE measurements for the two resonance frequencies display similar profiles and restricted range of values. The R_1 values exhibit the characteristic magnetic field strength dependence, since for proteins in the slow tumbling limit, R_1 values are highly dependent on the rates of motion occurring at the Larmor frequency (eq.1). The average $R_{1\alpha}^{500}$ over all Cα resonances is $2.07 \pm 0.34\ s^{-1}$, and $R_{1\alpha}^{600}$ is $1.48 \pm 0.17\ s^{-1}$. For the Cβ resonances, $R_{1\beta}^{500}$ is $2.26 \pm 0.32\ s^{-1}$ and $R_{1\beta}^{600}$ is $1.72 \pm 0.31\ s^{-1}$. The R_2 values in the slow tumbling limit are dominated by the zero frequency value of the spectral density, J(0) (eq.2), and are therefore similar at the different magnetic field strengths. The average $R_{2\alpha}^{500}$ is $23.4 \pm 6.3\ s^{-1}$ and $R_{2\alpha}^{600}$ is $20.7 \pm 5.1\ s^{-1}$. For Cβ, the average $R_{2\beta}^{500}$ is $20.2 \pm 6.7\ s^{-1}$ and $R_{2\beta}^{600}$ is $18.2 \pm 6.4\ s^{-1}$. The NOE values are highly sensitive to fast internal motion on the picosecond time-scale and the measurements are similar at the two field strengths. The average NOE_{α}^{500} is 1.15 ± 0.25 and NOE_{α}^{600} is 1.18 ± 0.15. For the Cβ resonances, NOE_{β}^{500} is 1.12 ± 0.29 and NOE_{β}^{600} is 1.25 ± 0.16.

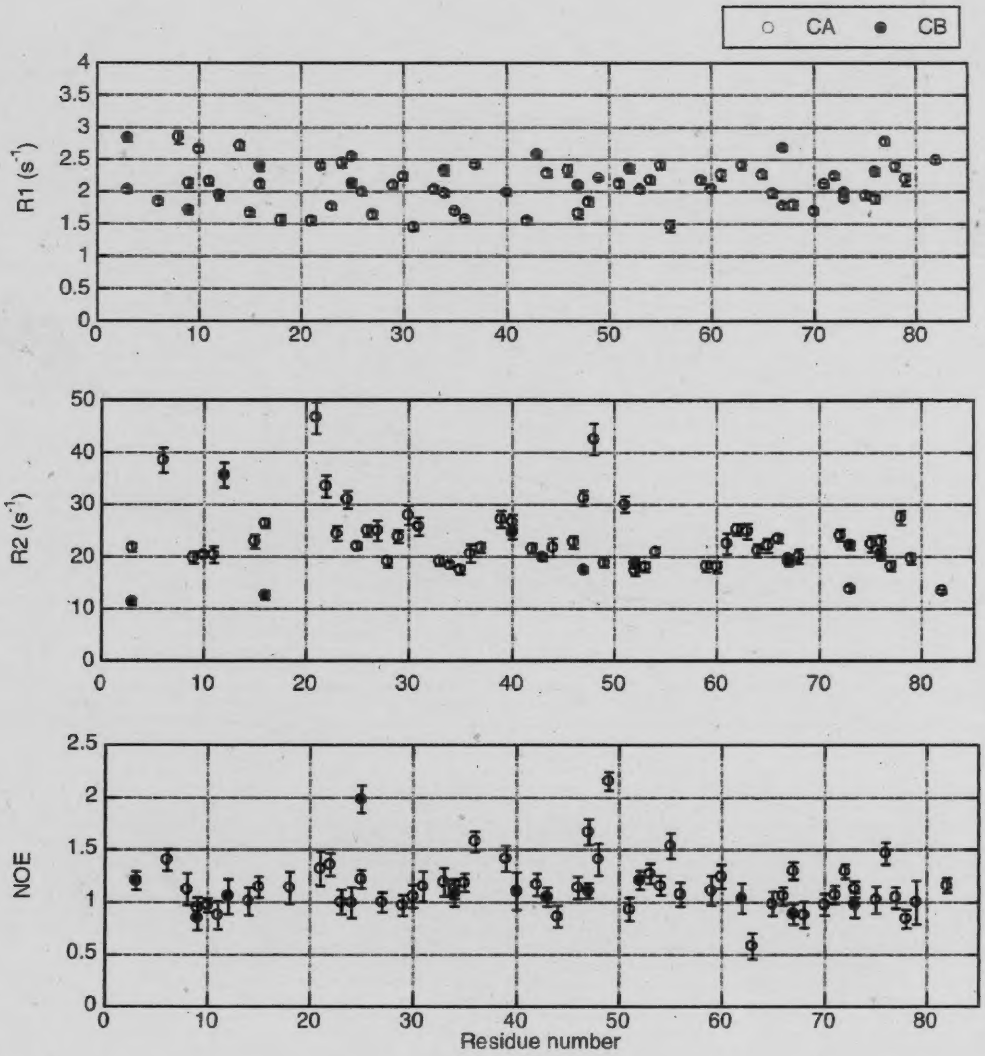


Figure 4.3: Plots of ^{13}C-R$_1$, ^{13}C-R$_2$, and {^{1}H}-^{13}C NOE at 500 MHz and 30°C with individual error bars plotted for Cα (open circles) and Cβ (filled circles).

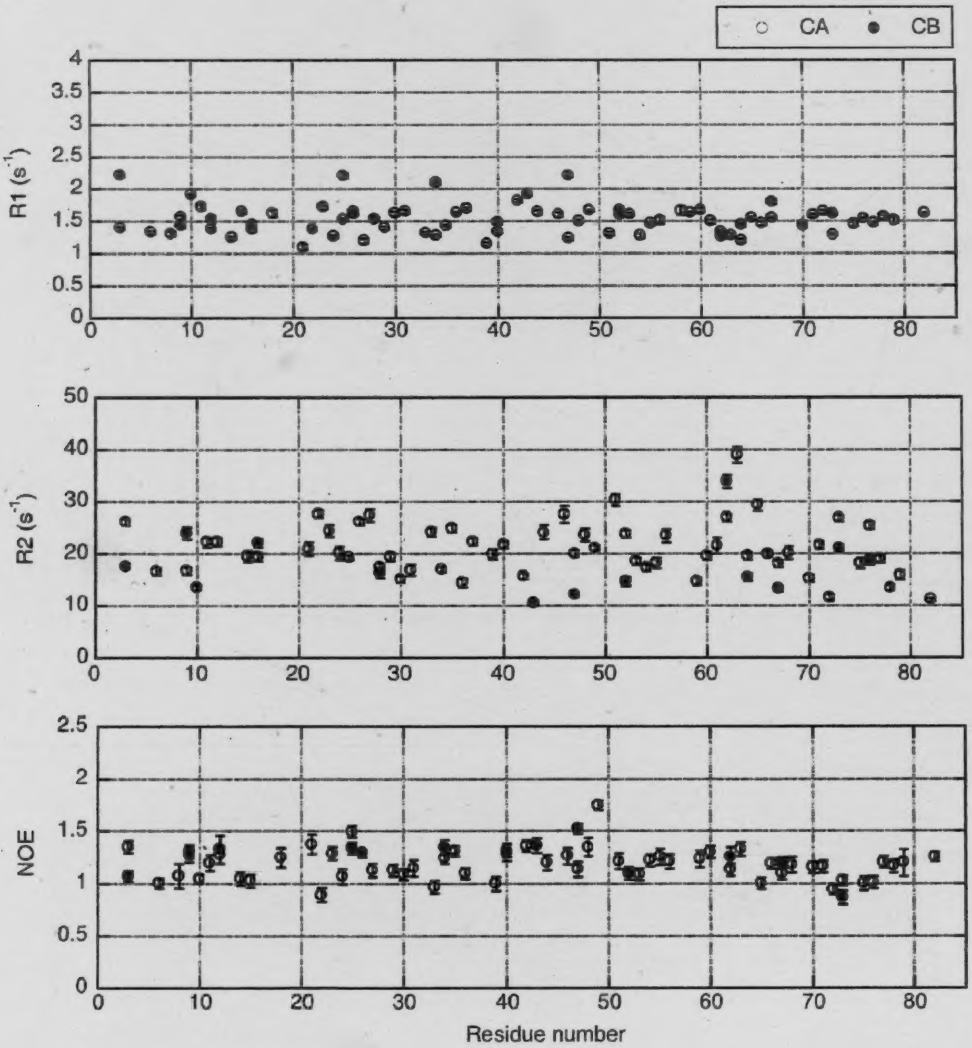


Figure 4.4: Plots of ^{13}C-R$_1$, ^{13}C-R$_2$, and {^{1}H}-^{13}C NOE at 600 MHz and 30°C with individual error bars plotted for Cα (open circles) and Cβ (filled circles).

The lower resolution of the spectra at 5°C allowed for analysis of fewer resonances because of increased spectral overlap. At 600 MHz, 39-43 Cα and 10-13 Cβ could be measured, while at 500 MHz 24-35 Cα and 12-13 Cβ measurements were used. The values of ^{13}C-R_1 and the NOE at 5°C are shown for the 500 MHz data in Figure 4.5 and the 600 MHz data in Figure 4.6. Again, at the two different resonance frequencies, the R_1 and NOE measurements display a similar profile and restricted range of values. The R_1 magnetic field strength dependence is obvious, where the average $R_{1\alpha}^{500}$ is 1.37±0.30 s^{-1} and $R_{1\alpha}^{600}$ is 0.98±0.36 s^{-1}. For the Cβ, $R_{1\beta}^{500}$ is 1.25±0.23 s^{-1} and $R_{1\beta}^{600}$ is 1.02±0.31 s^{-1}. The NOE measurements at 5°C are quite similar to their measurements at 30°C, with an average NOE_α^{500} of 1.25±0.39 and NOE_α^{600} of 1.17±0.21. The NOE_β^{500} is 1.36±0.35, while NOE_β^{600} is 1.27±0.31.

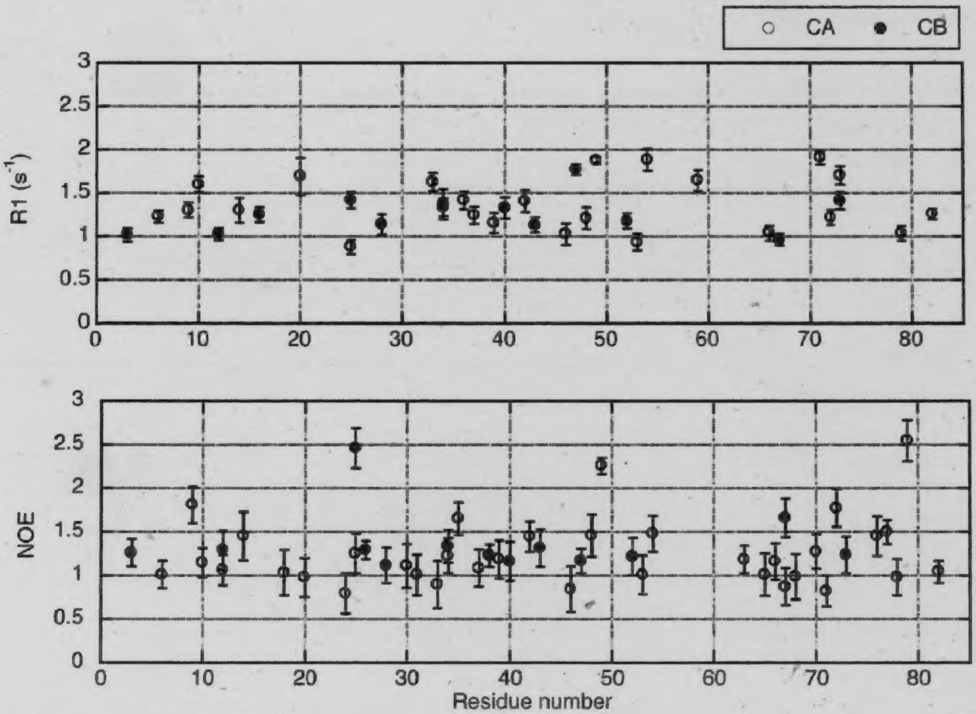


Figure 4.5: Plots of ^{13}C-R_1, and {^{1}H}-^{13}C NOE at 500 MHz and 5°C with individual error bars plotted for Cα (open circles) and Cβ (filled circles).

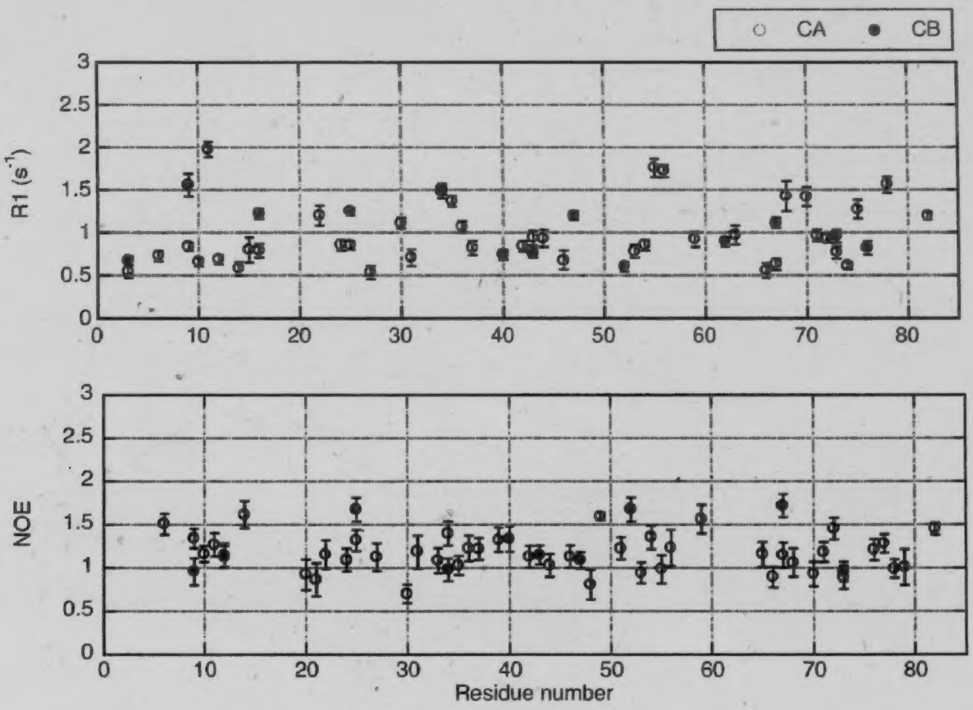


Figure 4.6: Plots of ^{13}C-R$_1$, and {^{1}H}-^{13}C NOE at 600 MHz and 5°C with individual error bars plotted for Cα (open circles) and Cβ (filled circles).

For the threonine and valine residues where both Cα and Cβ could be measured, the R$_1$, R$_2$, and NOE values of both resonances and their ratios (at 30°C and 600 MHz) are displayed in Tables 4.1, 4.2 and 4.3, respectively.

Table 4.1: Threonine and valine ^{13}C-R_1 measurements at 600 MHz and 30°C

Residue	Cα R_1 (s^{-1})	Cβ R_1 (s^{-1})	$R_{1\alpha}/R_{1\beta}$
Thr 3	1.41±0.02	2.21±0.02	0.64
Thr 9	1.56±0.03	1.44±0.04	1.09
Thr 12	1.38±0.02	1.53±0.05	0.90
Thr 16	1.37±0.02	1.45±0.02	0.94
Val 25	1.53±0.02	2.20±0.03	0.69
Thr 26	1.62±0.02	1.65±0.02	0.98
Thr 28	1.53±0.02	1.53±0.02	1.00
Val 34	1.28±0.02	2.09±0.03	0.61
Thr 40	1.34±0.02	1.49±0.03	0.90
Thr 47	1.23±0.03	2.20±0.02	0.56
Thr 52	1.60±0.03	1.67±0.02	0.96
Thr 62	1.34±0.02	1.27±0.03	1.06
Thr 64	1.20±0.02	1.45±0.03	0.83
Thr 67	1.55±0.02	1.80±0.03	0.86
Thr 73	1.30±0.02	1.61±0.02	0.81
Average	1.42	1.71	0.86
St. Dev.	0.14	0.32	0.16

Table 4.2: Threonine ^{13}C-R_2 measurements at 600 MHz and 30°C

Residue	Cα R_2 (s^{-1})	Cβ R_2 (s^{-1})	$R_{2\alpha}/R_{2\beta}$
Thr 3	26.10±0.70	17.55±0.55	1.49
Thr 9	16.73±0.67	23.88±1.19	0.70
Thr 16	19.39±0.77	21.97±0.78	0.88
Thr 28	17.43±0.67	16.27±1.03	1.07
Thr 47	19.99±0.65	12.05±0.45	1.66
Thr 52	23.74±0.71	14.61±0.84	1.63
Thr 62	26.84±0.79	33.82±1.26	0.79
Thr 64	19.45±0.72	15.46±0.78	1.26
Thr 67	18.04±0.62	13.18±0.51	1.37
Thr 73	26.70±0.60	21.05±0.68	1.27
Thr 76	25.28±0.66	18.63±0.76	1.36
Average	22.11	17.89	1.35
St. Dev.	3.81	6.29	0.44

Table 4.3: Threonine and valine NOE measurements at 600 MHz and 30°C

Residue	Cα NOE	Cβ NOE	NOEα/NOEβ
Thr 3	1.35±0.06	1.07±0.05	1.26
Thr 9	1.28±0.07	1.30±0.06	0.98
Thr 12	1.28±0.09	1.33±0.13	0.96
Val 25	1.50±0.06	1.34±0.05	1.12
Val 34	1.24±0.05	1.35±0.07	0.92
Thr 40	1.31±0.06	1.30±0.08	1.01
Thr 47	1.13±0.07	1.52±0.05	0.74
Thr 62	1.13±0.06	1.25±0.09	0.90
Thr 67	1.09±0.05	1.19±0.05	0.92
Thr 73	1.03±0.04	0.86±0.06	1.19
Average	1.23	1.25	1.00
St. Dev.	0.14	0.18	0.15

4.4.2 Estimation of the overall correlation time

The overall rotational correlation time (τ_c) of TmAFP in D_2O was calculated with a grid search to find the minimum in the squared difference measured and calculated R_2/R_1 ratios of the Cα resonances using an in-house written program (*38*). The calculation of τ_c was optimized using those residues that do not display obvious signs of relaxation active mobility using the criteria of Tjandra et al. (*39, 40*). The R_2/R_1 ratios of the selected residues should be dependent only on the global reorientation of the molecule, assuming negligible internal mobility and exchange. The global τ_c was then calculated on a per residue basis using the spectral density for isotropic rotational diffusion and the results averaged. A global τ_c of 4.5 ns at 600 MHz was obtained at 30°C. This value is slightly higher than the τ_c of 4 ns obtained from the ^{15}N relaxation data (*24*), but is fully consistent with the increase in viscosity (η) of D_2O compared to H_2O. The value of η^{H2O} at 30°C is 0.797 cP, while η^{D2O}=0.976 cP (*41*), with a ratio of η^{H2O}/η^{D2O}=0.82. The ratio $\tau_c^{H2O}/\tau_c^{D2O}$=0.89, a value comparable to the ratio of the viscosities, indicating that the increase is attributable to the D_2O solution.

4.5 Discussion

In TmAFP we found that the τ_c^{D2O} calculated using the backbone CαH resonances was nearly equivalent to that measured previously from ^{15}N relaxation of NH vectors, allowing for the increased viscosity of D_2O. This indicates that the data from the two studies can be compared directly and that the CαH bond vectors are similarly rigid. Overall, it appears that the main chain CαH vectors behave in an analogous fashion to the restrictive motions of the backbone amide vectors. By comparing data from the CαH to the CβH at 30°C, it is clear that in general the average values of R_1, R_2 or NOE are not significantly different and are within the error limits over the whole range of measurements. In particular, for the threonine and valine residues that have both the CαH and CβH values measured from the same side chain, the calculated ratios shown in the tables indicate an overall restriction of motion for the CβH with respect to the CαH. Despite the fact that all of the threonine and valine side chains are presented on the surface of the protein, for most of the residues we do not observe increasing flexibility in the measurements towards the end of the side chain. It appears that for the threonine residues in TmAFP, the Cβ resonances are as rigid as the backbone and can be considered functionally as part of the backbone. The dominant restriction to this motion may be the favourable hydrogen bonding interactions that can be formed between the threonine side chain and the rigid backbone.

Theoretical plots for the R_1, R_2 and NOE as a function of γ_0 at various internal correlation times are shown in Figures 4.7, 4.8 and 4.9, respectively. These were calculated using the spectral density function in equation 5, where τ_c is assumed to be 4.5 ns. Although from the ^{15}N relaxation data analysis we determined that TmAFP undergoes axially symmetric anisotropic rotation, here we do not include anisotropic overall reorientation in the calculations. All calculations were performed using a τ_c of 4.5 ns, the 30°C correlation time. At this temperature, the effects of axially symmetric rotation are not as pronounced as at 5°C. However, the conclusions at 30°C are also valid for the data at 5°C, as on examination of the raw R_1 and NOE data presented in Figures 4.5 and 4.6, it is clear that the Cβ values are again in the same range as the Cα values. Therefore the discussion that follows can be extended to the data at both temperatures.

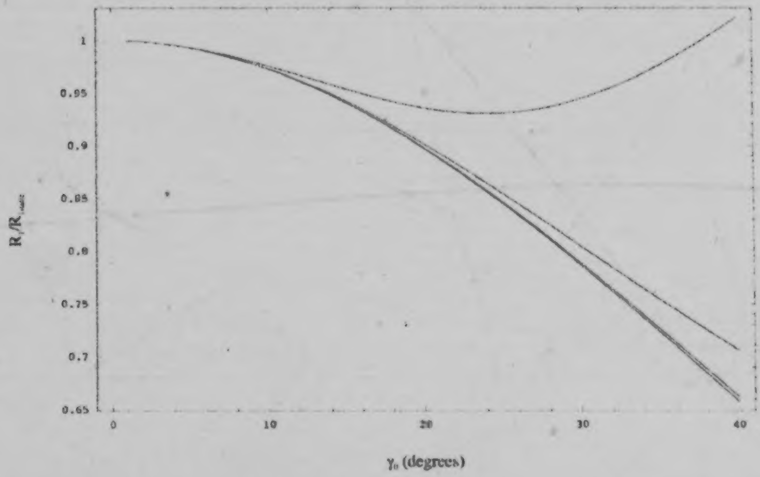


Figure 4.7: Calculated graph of ^{13}C $R_1/R_{1static}$ vs γ_0 at four internal correlation times ($\tau_i=10^{-9}$, 10^{-10}, 10^{-11}, and 10^{-12}, from top to bottom), where $R_{1static}$ is the longitudinal relaxation rate in the absence of internal motion.

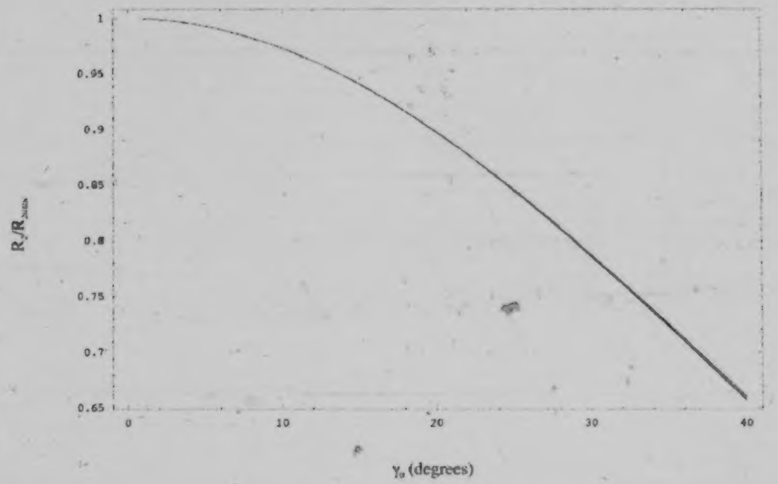


Figure 4.8: Calculated graph of ^{13}C $R_2/R_{2static}$ vs γ_0 at four internal correlation times ($\tau_i=10^{-9}$, 10^{-10}, 10^{-11}, and 10^{-12}, from bottom to top), where $R_{2static}$ is the transverse relaxation rate in the absence of internal motion.

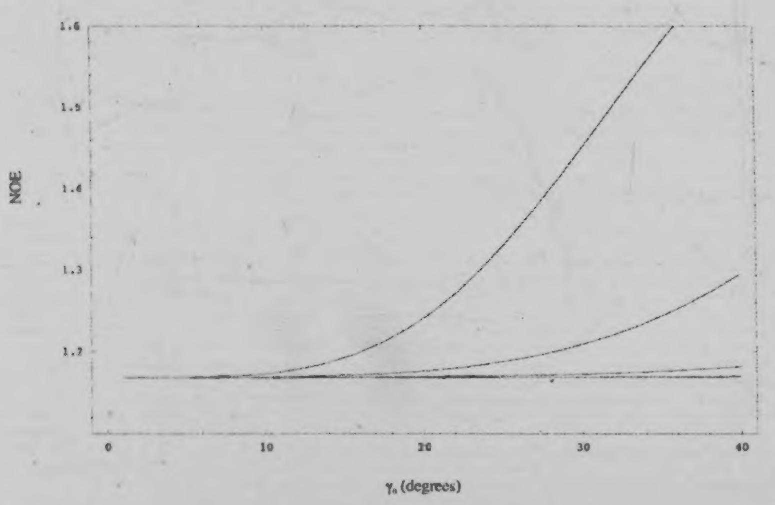


Figure 4.9: Calculated graph of $\{^1H\}$-^{13}C NOE vs γ_0 at four internal correlation times (τ_i=10^{-9}, 10^{-10}, 10^{-11}, and 10^{-12}, from top to bottom).

The results for $R_1/R_{1static}$, $R_2/R_{2static}$, and NOE as a function of γ_0 and corresponding to the values of D_0 and D_i specified are summarized in Figure 4.7, 4.8 and 4.9, respectively. The values for each parameter are calculated out to γ=40°. These results indicate that the relaxation parameters display very little dependence for smaller angles of displacement about the χ_1 angle. For the R_1, R_2 and NOE measurements of TmAFP, the values obtained indicate that the threonine side chains undergo restricted motions, corresponding to $\gamma_0 \leq 25°$.

4.6 Conclusions

Using a combination of experimental heteronuclear relaxation rates at ^{13}C natural abundance and rotational diffusion model calculations, we have demonstrated that the threonine side chains of the *Tenebrio molitor* antifreeze protein do not undergo significant motion about the χ_1 dihedral angle. This result agrees with our previous study of χ_1 angles by examination of $^3J_{\alpha\beta}$ coupling constants. In that study we observed that most of the threonine residues (including all of the ice-binding threonines) adopted specific rotameric conformations and did not experience a significant amount of rotameric averaging. Furthermore, by using heteronuclear relaxation measurements, we have identified that the restriction of motion occurs on the pico- to nanosecond time scale of bond vector and librational motions. It was not possible to make this conclusion on the basis of motionally averaged J couplings, which can arise from a combination of fast and slow time scale motions. The rigidity of the side chains in this small, β-helical protein is likely important for its ice binding activity, as it may reduce the entropic barrier for binding to a flat surface.

4.7 References

1. Kay, L.E. (1998) *Biochem. Cell Biol.*, **76**, 145-152.

2. Kay, L.E., Muhandiram, R., Wolf, G., Shoelson, S.E., and Forman-Kay, J.D. (1998) *Nat. Struct. Biol.*, **5**, 156-163.

3. Yamazaki, T., Muhandiram, R., and Kay, L.E. (1994) *J. Am. Chem. Soc.*, **116**, 8266-8278.

4. Engelke, J. and Rüterjans, H. (1998) *J. Biomol. NMR*, **11**, 165-183.

5. Wand, A.J., Urbauer, J.L., McEvoy, R.P., and Bieber, R.J. (1996) *Biochemistry*, **35**, 6116-6125.

6. LeMaster, D.M. and Kushlan, D.M. (1996) *J. Am. Chem. Soc.*, **118**, 9255-9264.

7. Muhandiram, R., Yamazaki, T., Sykes, B.D., and Kay, L.E. (1995) *J. Am. Chem. Soc.*, **117**, 11536-11544.

8. Yang, D., Mittermaier, A., Mok, Y.K., and Kay, L.E. (1998) *J. Mol. Biol.*, **276**, 939-954.

9. Nirmala, N.R. and Wagner, G. (1988) *J. Am. Chem. Soc.*, **110**, 7557-7558.

10. Dellwo, M.J. and Wand, A.J. (1989) *J. Am. Chem. Soc.*, **111**, 4571-4578.

11. Palmer, A.G., Rance, M., and Wright, P.E. (1991) *J. Am. Chem. Soc.*, **113**, 4371-4380.

12. Mispelter, J., Lefevre, C., Adjadj, E., Quiniou, E., and Favaudon, V. (1995) *J. Biomol. NMR*, **5**, 233-244.

13. Sklenar, V., Torchia, D., and Bax, A. (1987) *J. Magn. Reson.*, **73**, 375-379.

14. Kay, L.E. and Torchia, D.A. (1991) *J. Magn. Reson.*, **95**, 536-547.

15. Palmer, A.G., Wright, P.E., and Rance, M. (1991) *Chem. Phys. Lett.*, **185**, 41-46.

16. Guenneugues, M., Gilquin, B., Wolff, N., Ménez, A., and Zinn-Justin, S. (1999) *J. Biomol. NMR*, **14**, 47-66.

17. Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4546-4559.

18. Lipari, G. and Szabo, A. (1982) *J. Am. Chem. Soc.*, **104**, 4559-4570.

19. Clore, G.M., Driscoll, P.C., Wingfield, P.T., and Gronenborn, A.M. (1990) *Biochemistry*, **29**, 7387-7401.

20. Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C., and Gronenborn, A.M. (1990) *J. Am. Chem. Soc.*, **112**, 4989-4991.

21. Lipari, G. and Szabo, A. (1980) *J. Biophys.*, **30**, 489-506.

22. Lipari, G. and Szabo, A. (1981) *J. Chem. Phys.*, **75**, 2971-2976.

23. London, R.E. (1980) In *Magnetic Resonance in Biology* (Cohen, J.S., ed.) Vol. 1, Wiley, New York, pp. 1-69.

24. Daley, M.E., Spyracopoulos, L., Jia, Z., Davies, P.L., and Sykes, B.D. (2002) *Biochemistry*, **41**, 5515-5525.

25. Liou, Y.-C., Daley, M.E., Graham, L.A., Kay, C.M., Walker, V.K., Sykes, B.D., and Davies, P.L. (2000) *Protein Expr. Purif.*, **19**, 148-157.

26. Marshall, C.B., Daley, M.E., Graham, L.A., Sykes, B.D., and Davies, P.L. (2002) *FEBS Lett.*, **529**, 261-267.

27. Daley, M.E. and Sykes, B.D. (2003) *Protein Sci.*, **12**, 1323-1331.

28. Karplus, M. (1959) *J. Chem. Phys.*, **30**, 11-15.

29. Karplus, M. (1963) *J. Am. Chem. Soc.*, **85**, 2870-2871.

30. Clore, G.M. and Gronenborn, A.M. (1989) *Crit. Rev. Biochem. Mol. Biol.*, **24**, 479-557.

31. Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994) *Biochemistry*, **33**, 5984-6003.

32. Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biomol. NMR*, **6**, 277-293.

33. Johnson, B.A. and Blevins, R.A. (1994) *J. Biomol. NMR*, **4**, 603-614.

34. London, R.E. and Avitabile, J. (1978) *J. Am. Chem. Soc.*, **100**, 7159-7165.

35. Wittebort, R.J. and Szabo, A. (1978) *J. Chem. Phys.*, **69**, 1722-1736.

36. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Clarendon Press, Oxford, U.K.

37. Wolfram, S. (2002) *Mathematica 4.2*. Wolfram Research Inc. (http://www.wolfram.com)

38. Gagné, S.M., Tsuda, S., Spyracopoulos, L., Kay, L.E., and Sykes, B.D. (1998) *J. Mol. Biol.*, **278**, 667-686.

39. Tjandra, N., Feller, S.E., Pastor, R.W., and Bax, A. (1995) *J. Am. Chem. Soc.*, **117**, 12562-12566.

40. Tjandra, N., Wingfield, P., Stahl, S., and Bax, A. (1996) *J. Biomol. NMR*, **8**, 273-284.

41. Cho, C.H., Urquidi, J., Singh, S., and Robinson, G.W. (1999) *J. Phys. Chem. B*, **103**, 1991-1994.

CHAPTER 5: The Temperature Dependence of Amide Proton Chemical Shifts

5.1 Introduction

The chemical shift of a nucleus has been recognized for many years as a sensitive indicator of molecular conformation, composition, and environment. In proteins, ^{1}H, ^{13}C, and ^{15}N chemical shifts are all used to provide information about the environment of a particular nucleus, but the complexity of biomolecular systems has made a complete understanding of the basis for protein chemical shifts elusive. Proteins have long been noted to have a wide dispersion of chemical shifts relative to random coil reference data (*1*). In particular, chemical shifts are exquisitely sensitive to the protein secondary structure and the noted trends are used to deduce preliminary structural information (*2*). Upfield shifts for nuclei in α-helical conformations and downfield shifts for nuclei in β-strand and extended conformations are seen for α- and amide ^{1}H, α- and carbonyl ^{13}C, and amide ^{15}N chemical shifts (*2*). The relationship between secondary structure and amide chemical shifts is probably the least well-understood of these.

The dependence of the chemical shift of amide proton resonances on temperature however, has been known for many years (*3*). The general trend is that the amide protons experience an upfield shift upon warming, described as a negative temperature coefficient. This effect has been rationalized as a weakening of hydrogen bonding with increased temperature (*4*). In a hydrogen-bonded amide, the carbonyl hydrogen bond partner causes a downfield shift of the amide resonance. With the lengthening of the hydrogen bond caused by increased temperature, the amide proton is less downfield shifted, resulting in a relative upfield shift. Therefore, solvent exposed amides will experience a more rapid upfield shift upon warming owing to the decrease in intermolecular versus intramolecular hydrogen bonding (*5*). This has led to the widely used practice of attributing temperature gradient ($\Delta\delta/\Delta T$) values less negative than ~ –4 to –4.5 ppb/°C to the sequestration from solvent of the amide proton. However, this appears to be an oversimplified approach, and other factors including the chemical shift deviation (CSD) of the resonances from random coil values and the degree of structuring of the protein or peptide of interest (*5*) must be included in the analysis. For example, Baxter and Williamson (*6*) have shown that taking exchange data into consideration with temperature gradients increases the reliability of determining hydrogen bonding, while Andersen et al. (*5*) have demonstrated the utility of correlation plots of NH-CSD versus $\Delta\delta/\Delta T$. Although it has been noted that hydrogen bonded amides in rigid proteins are usually characterized by $\Delta\delta/\Delta T$ values of -2.0±1.4 ppb/°C and exposed amides are more negative, numerous exceptions exist (*5*).

We have applied this analysis to the amide protons of the β-helical antifreeze protein from the beetle *Tenebrio molitor* (TmAFP). We have previously studied the ^{15}N backbone dynamics of TmAFP and have shown that it is a rigid protein that displays restricted internal motions throughout at both 30 and 5°C (*7*). In the process of assigning the ^{15}N HSQC spectrum at 5°C for the analysis of ^{15}N relaxation, we noticed some interesting patterns of the temperature gradients and chemical shifts with decreasing temperature and decided to pursue these observations further.

TmAFP is a small (8.4 kDa), highly disulfide-bonded right handed β-helix (Figure 5.1).

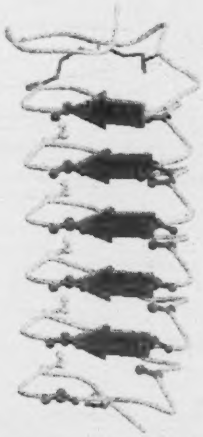


Figure 5.1: TmAFP ribbon structure (PDB 1EZG; *8*) with ice-binding threonines displayed in ball-and-stick representation and coloured blue. The eight disulfide bonds are coloured red, and the inward-facing serine and alanine residues that line the inside of the coil are also displayed in ball-and-stick representation and coloured light blue and green, respectively. Figure prepared using Molscript (*9*) and Raster3D (*10*).

Both the solution (*7*) and crystal (*8*) structures have been solved and have provided complementary observations. From the NMR studies, information about the rigidity of the backbone and side chain conformations (*11*) has been gained, while the crystal structure provided detailed insight into the unique hydrogen bonding network with external water molecules that mimic a layer of ice (*8*). This study of the amide proton chemical shifts and temperature gradients supports the previous conclusions that the ice-binding face is rigid, and provides more information on the unique structure and behaviour of this unusual protein.

5.2 Experimental Procedures

5.2.1 NMR spectroscopy

The ^{15}N-labeled TmAFP sample used was that previously described (*7*). The sample was prepared for NMR spectroscopy by dissolving the lyophilized protein in 90% H_2O / 10% D_2O containing 0.1 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid). The final protein concentration was approximately 0.4 mM, and the pH was adjusted to 5.5 with microliter aliquots of 100 mM NaOD or DCl as required. A series of ^{1}H-^{15}N HSQC spectra for the temperature range of 30°C to 5°C were collected in 5°C decrements. The spectra were acquired under identical conditions using a Varian Unity 600 MHz spectrometer equipped with a 5 mm triple resonance probe and *z*-axis pulsed field gradients. After each temperature change, the sample was allowed to equilibrate for 45 minutes. Each spectrum was correctly referenced to DSS by collecting a ^{1}H-1D spectrum at each temperature using the same ^{1}H spectral width as for the HSQC. The spectral width for ^{1}H was 7000.35 Hz and for ^{15}N was 1897.89 Hz. The number of real data points acquired for ^{1}H and ^{15}N were 436 and 128, respectively, with a total of 64 transients collected. The NMR data was processed and displayed using VNMR 5.3 software on a Sun workstation. The F_2 (^{1}H) dimension was multiplied by a 60°-shifted sine-bell function and the F_1 dimension was multiplied by a 90°-shifted squared sine-bell function before Fourier transformation.

5.3 Results and Discussion

The ^{15}N HSQC spectra at the six temperatures are shown in Figure 5.2. As the sample is cooled to 5°C, the amides display the typical downfield shift.

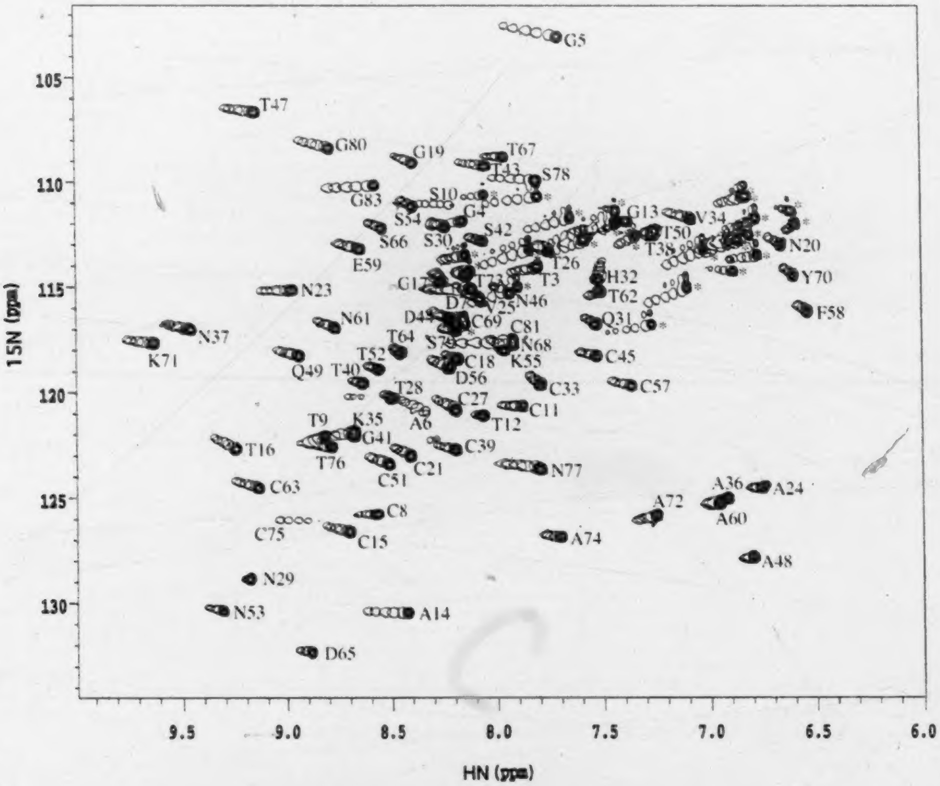


Figure 5.2: Plot of the ^{15}N HSQC spectra displayed as a function of temperature. The 30°C spectrum is plotted with 10 contours, the remaining spectra are plotted with 1 contour at 5°C intervals down to 5°C. The amide resonances are labeled with their assignments, those labeled with a * are Asn and Gln side chain NHs.

Upon closer inspection it was noted that some residues moved very little or not at all. For example, note T28, T40, T52 and T64 near the middle of the spectrum or N29 in the bottom left corner. Therefore, we plotted the changes in chemical shift of the individual residues on the TmAFP protein surface (Figure 5.3) to determine if any overall pattern emerged.

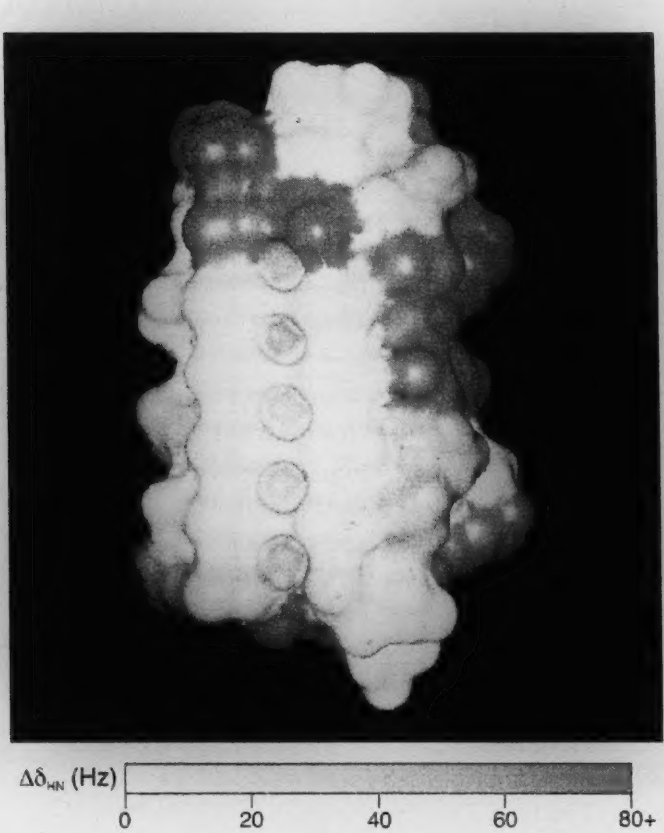


Figure 5.3: The change in amide chemical shifts from 30°C to 5°C mapped on the TmAFP surface (PDB 1EZG). Figure prepared with InsightII.

It immediately became obvious that the threonine residues on the ice binding face of the protein experience very little change in chemical shift as the temperature is lowered. The behaviour of the surrounding residues varies much more widely. Some residues such as N29 also change little, while others, including A14, T16 and other residues adjacent to the threonines do change noticeably. For example, the cysteine residues forming the TCT motif of the ice binding face have much more negative temperature gradients than the threonines they separate.

Table 5.1: NH Shift Deviations and Temperature Gradients for TmAFP

Residue	NH shift deviation (ppm)	NH Δδ/ΔT (ppb/°C)
N68	-0.36	-10.64
C81	-0.30	-9.92
G5	-0.51	-8.44
S78	-0.42	-8.04
C75	0.52	-7.88
G83	0.34	-7.12
A14	0.24	-6.72
D7	-0.27	-6.36
K35	0.49	-6.24
S10	-0.08	-5.28
G4	-0.16	-5.12
A6	0.07	-4.80
T3	-0.60	-4.76
N23	0.53	-4.68
T47	0.95	-4.64
G80	0.47	-4.64
K71	1.38	-4.16
T43	-0.16	-4.08
C8	0.16	-4.08
N37	0.99	-3.88
T16	1.02	-3.52
V25	-0.07	-3.48
C15	0.30	-3.44
N77	-0.62	-3.40
C63	0.72	-3.28
V34	-1.07	-3.16
C69	-0.26	-3.04
C27	-0.23	-3.00
C39	-0.23	-2.84
A72	-1.04	-2.76
T12	-0.22	-2.76
C51	0.09	-2.76
Q49	0.59	-2.72
C57	-1.06	-2.68
T67	-0.28	-2.68
C21	-0.03	-2.56
D56	-0.27	-2.48
N61	0.27	-2.40
C45	-0.91	-2.40
N53	0.80	-2.28
G19	0.02	-2.24
N46	-0.56	-2.20
D44	-0.28	-1.96
A74	-0.61	-1.80
T76	0.53	-1.76
E59	0.30	-1.72
G13	-1.01	-1.68
C11	-0.54	-1.68
C18	-0.27	-1.56
N20	-1.85	-1.52
S42	-0.31	-1.52

Table 5.1
(continued)

Residue	NH shift deviation (ppm)	NH Δδ/ΔT (ppb/°C)
S66	0.17	-1.48
S79	-0.19	-1.40
S30	-0.14	-1.40
S54	0.01	-1.40
D65	0.37	-1.40
K55	-0.34	-1.32
A36	-1.41	-1.24
T26	-0.52	-1.24
T62	-0.76	-1.20
Q31	-0.86	-1.08
Y70	-1.65	-1.00
A60	-1.38	-1.00
A24	-1.58	-1.00
T73	-0.13	-0.92
T52	0.28	-0.84
A48	-1.55	-0.68
F58	-1.84	-0.64
T40	0.35	-0.60
C33	-0.67	-0.48
T28	0.20	-0.40
G41	0.26	-0.40
T64	0.16	-0.36
T38	-1.04	-0.20
T50	-1.04	-0.20
G17	-0.18	-0.20
N29	0.63	-0.12
T9	0.53	0.64
H32	-1.02	1.72

The temperature gradients (ppb/°C) for each residue are shown in Table 5.1, which displays both the CSD and Δδ/ΔT and is sorted by the value of the temperature gradient. To illustrate the previous result, the temperature gradients for the threonine and cysteine residues in the TCT repeats of the four central loops were separately averaged. For the left hand rank of threonines (T28, T40, T52, T64) the average is –0.55 ppb/°C, and for the right hand rank of threonines (T26, T38, T50, T62) it is –0.71 ppb/°C. In contrast, for the cysteines between these threonines (C27, C39, C51, C63), which form intraloop disulfide bonds, the average was –2.97 ppb/°C. We also observe that the inward-facing alanine and serine residues that line the right and left sides of the protein core respectively (Figure 5.1), also cluster in the temperature gradient table and have less negative values. The average for the alanines (A24, A36, A48, A60) is –0.98 ppb/°C and for the serines (S30, S42, S54, S66) is –1.45 ppb/°C.

Using the general practice of assuming that $\Delta\delta/\Delta T$ values less negative than –4 ppb/°C indicate NH sequestration from solvent would suggest that all of these residues are therefore participating in hydrogen bonds and are protected. However, by plotting the data in a correlation plot of amide proton CSD versus $\Delta\delta/\Delta T$ (*5*) a very different conclusion is reached. The CSD values are calculated at the lowest temperature of the experiment as $\delta_{obs} - \delta_{ref}$, so that upfield shifts relative to the random coil chemical shift are recorded as negative CSD values. The reference random coil values used are those of Andersen et al. (*5*), with the specified correction for temperature of –7.6 ppb/°C (*12*) to adjust the reference values to 5°C. As devised by Andersen et al. (*5*), the plot is formatted with NH shift deviation on the x-axis with downfield NH resonances to the left as in an NMR spectrum. The temperature gradient (y) axis is oriented so that amides which experience larger upfield shifts upon warming are at the top of the graph. Figure 5.4 shows the data for TmAFP plotted in this fashion. In this plot, points appearing toward the lower left represent downfield NHs which do not shift rapidly; these are likely to be intramolecularly hydrogen-bonded sites.

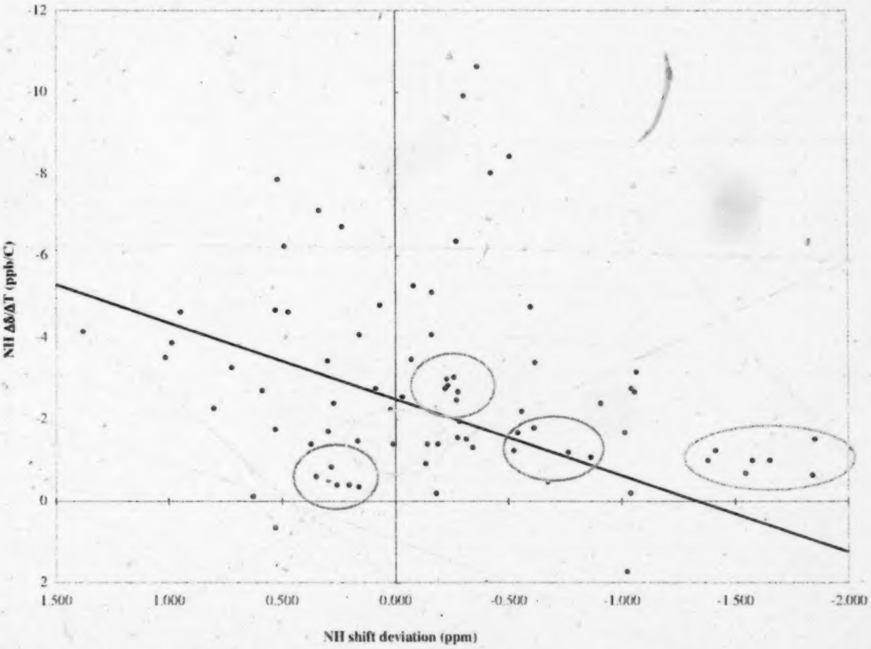


Figure 5.4: Amide CSD – $\Delta\delta/\Delta T$ correlation for the NHs in TmAFP. The line corresponds to $\Delta\delta/\Delta T = -2.97 - (\mathrm{CSD} \bullet 2.19)$ as used in reference *5*. The circles are coloured for residues as in Figure 5.1 and discussed in the text.

The cutoff line used corresponds to one used by Andersen et al. (*5*) to provide "superior discrimination" between exchange-protected and exposed amide protons. With this plot, we re-examine the results for the previously discussed residues. First, the TCT ice-binding motif, for which there is little change in the chemical shift of the threonines over the 25°C temperature range measured. An interesting observation regarding the CSD for these threonines is noted; although both ranks of threonine have similar values of $\Delta\delta/\Delta T$, they experience different directions in chemical shift relative to the random coil values. The threonines in the left rank (T28, T40, T52, T64) have an average CSD of 0.25 ppm downfield and are well within the area of the plot corresponding to high potential hydrogen bond formation as marked by the leftmost blue circle. However, the threonines in the right hand rank experience an average upfield shift of –0.84 ppm relative to random coil values, putting them on the edge of the designated line and marked as the right hand blue circle, at which point conclusions regarding the hydrogen bonding state of these residues cannot necessarily be made. The cysteine residues between the two ranks of threonine also do not appear to be hydrogen bonded; though they have a –2.97 ppb/°C temperature gradient, they also tend to cluster around a CSD value of approximately –0.2 ppm as designated by the red circle.

The amides of the serine residues have values quite close to the random coil values, therefore the CSD is small and with the –1.45 ppb/°C $\Delta\delta/\Delta T$ value as shown by the light blue circle in the plot, they also appear to be participating in intramolecular hydrogen bonds. In the TmAFP crystal structure (*8*), the serine NH forms an intraloop hydrogen bond to stabilize the tight γ-turn. The alanine residues, on the other hand, which have an amide $\Delta\delta/\Delta T$ value of –0.98 ppb/°C, are shifted very far upfield, approximately –1.5 ppm from the random coil values and designated by the green circle. This may be explained by the environment surrounding these residues, which is a constrained β-turn requiring very restrictive backbone dihedral angles. Accordingly, from Figure 5.4, they appear not to participate in intramolecular hydrogen bonds. A similar observation for a serine with a temperature gradient of approximately –1.5 ppb/°C and a strong upfield shift in a non-hydrogen bonded amide was made in ubiquitin (*13*). In the crystal structure of TmAFP, the alanine-containing channel also contains five internal water molecules spaced regularly between the loops (*8*). These internal waters form hydrogen bonds to three residues each, including the alanine NH. Hydrogen bonding to internal water molecules is likely not as stable as intramolecular hydrogen bonding to other atoms in the protein, since they could be labile, allowing rearrangement of the hydrogen bonding network. This may contribute to the behaviour of these alanines in the temperature coefficient and CSD correlation plot.

5.4 Conclusions

Overall, a significant portion of the residues in TmAFP appear to be intramolecularly hydrogen bonded. This result is in agreement with expectations for parallel β-sheet structure, which should contain numerous interstrand hydrogen bonds. Several interesting deviations were noted, particularly for the alanine residues. These results may be explained by the unique environment of the alanines in this highly constrained structure as well as by the participation of internal water molecules in the hydrogen bonding network. The threonine residues of the ice-binding face likely form intramolecular, interloop hydrogen bonds to help stabilize the flat surface, although there appear to be slight differences manifest as chemical shift changes between the two ranks of threonine residues.

5.5 References

1. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*. John Wiley & Sons, New York.

2. Wishart, D.S., Sykes, B.D., and Richards, F.M. (1991) *J. Mol. Biol.*, **222**, 311-333.

3. Ohnishi, M. and Urry, D.W. (1969) *Biochem. Biophys. Res. Commun.*, **36**, 194-202.

4. Llinás, M. and Klein, M.P. (1975) *J. Am. Chem. Soc.*, **97**, 4731-4737.

5. Andersen, N.H., Neidigh, J.W., Harris, S.M., Lee, G.M., Liu, Z., and Tong, H. (1997) *J. Am. Chem. Soc.*, **119**, 8547-8561.

6. Baxter, N.J. and Williamson, M.P. (1997) *J. Biomol. NMR*, **9**, 359-369.

7. Daley, M.E., Spyracopoulos, L., Jia, Z., Davies, P.L., and Sykes, B.D. (2002) *Biochemistry,* **41,** 5515-5525.

8. Liou, Y.-C., Tocilj, A., Davies, P.L., and Jia, Z. (2000) *Nature*, **406**, 322-324.

9. Kraulis, P.J. (1991) *J. Appl. Crystallogr.*, **24**, 946-950.

10. Merritt, E.A. and Bacon, D.J. (1997) *Methods Enzymol.*, **277**, 505-524.

11. Daley, M.E. and Sykes B.D. (2003) *Protein Sci.*, **12**, 1323-1331.

12. Merutka, G., Dyson, H.J., and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14-24.

13. Cierpicki, T., Zhukov, I., Byrd, R.A., and Otlewski, J. (2002) *J. Magn. Reson.*, **157**, 178-180.

CHAPTER 6: General Conclusions

The amazing structural diversity of AFPs and the question of how they attain the same function intrigues both evolutionary and structural biologists. The multiple evolutionary origins of AFPs have been confirmed by the identification of the progenitors of several fish antifreezes (*1*). These precursors are divergent proteins of quite unrelated functions. As well, the distribution of different types of AFPs does not coincide with the evolutionary relationships of the fish (*2*). For example, Type II AFP, a homologue of the carbohydrate recognition domain of Ca^{2+}-dependent lectins, is found in three distantly related fishes (sea raven, herring, and smelt), whereas two very closely related fishes (shorthorn sculpin and longhorn sculpin) from the same genus produce completely unrelated Type I and Type IV AFP (*3*). This implies that the selective force of sea-level glaciation is a recent phenomenon and that the threat of freezing that forced the fish to adapt occurred after the present sub-orders, families and genera were established. Therefore a number of very dissimilar proteins were adapted to the function of binding ice (*2*). The insect AFPs provide a remarkable example of convergent evolution since they are both β-helices, but are not homologous. TmAFP is right-handed with 12 amino acid residues per loop, while sbwAFP is left-handed and has 15 amino acids per loop. The two proteins have completely different disulfide bonding patterns as well, yet their ice-binding surfaces both contain a virtually superimposable array of threonine residues.

However, with this knowledge we still have not answered the question of whether all these diverse AFPs are operating by the same mechanism. The ability to bind ice suggests that these proteins must have some shared properties that enable them to perform their common function, however these may be obscured by the large structural differences. Typically, AFPs have been divided into two broad classes, those that are repetitive and those that are not (*1*) (Table 1.1 summarizes the repeat patterns). The Type I and insect AFPs provide the clearest examples of repetitive sequence and structure and are consistent with a lattice matching model for ice binding. The non-repetitive Type II and Type III AFPs however, do share one aspect of their development as both of these have evolutionary precursors that interact with carbohydrates. For Type II AFP, it is the previously mentioned carbohydrate recognition domain of C-type lectins, while for Type III AFP it is sialic acid synthase (*4*). Despite these differences in structure and evolutionary origin, I will argue that all AFPs do function by a similar mechanism directly related to protein stability and dynamics.

As has already been mentioned, TmAFP is the smallest β-helix yet seen in proteins, with only 12 amino acids per loop. Because of this, it has no hydrophobic core and is held together by eight disulfide bonds and extensive hydrogen bonding interactions. This causes difficulty in folding the recombinantly produced protein, which led us to examine the folded state of the protein by NMR (see Appendices B-E). However in my experience handling this protein, once TmAFP is in the correctly folded conformation it is extremely stable. For instance, the unlabeled Tm 2-14 used for experiments described in Chapters 2, 3, and 4 and Appendices B and C is the same sample I first received in November 1998 and the quality of the spectra obtained has been exceptional at all times over these five years. This leads me to believe that the β-helix is an intrinsically stable fold and provides an ideal scaffold on which to build an ice binding face.

The ^{15}N and ^{13}C backbone dynamics experiments indicate that TmAFP undergoes restricted internal mobility throughout the backbone at both 30°C and 5°C. The backbone dynamics of the sbwAFP have also been recently studied (*5*). The sbwAFP is observed to become more rigid as the temperature is lowered toward the physiological temperature. This study supports the observation with TmAFP that the β-helix is a very stable structure and that the protein rigidity may serve to provide a better lattice match to the flat ice surface that it binds. The rigid and repetitive structural elements can ensure the correct positioning of the required protein constituents. We also suggest that this hydrophobic binding site and its rigidity would reduce the entropic penalty for binding to ice.

The conservation of hydrophobic interactions in the ice-binding interface has been an interesting development in the AFP mechanism. The mutations of the threonine residues in Type I AFP to serine and valine showed that neither the hydroxyl group nor the methyl group alone was sufficient to confer full ice binding activity. The valine mutants had greater activity, leading to the first suggestions that hydrophobic contacts were more important (*6-9*). Type I AFP has undergone a further revision of its ice binding face with the realization that conserved alanine residues adjacent to the threonine residues were equally responsible for conferring ice binding activity (*10*). In combination with work identifying the ice-binding surfaces of Type III AFP (*11*), sbwAFP (*12*), and TmAFP (*13*; see Appendix C), a few features common to the ice-binding faces can be noted. First, side chain mutations that either lengthen the side chain and sterically block the AFP-ice interaction or that shorten the side chain leaving a cavity in the ice-binding surface are deleterious. Second, these mutations are tolerated elsewhere on the protein. Third, hydrogen bond formation between polar side chains and ice may occur, but is limited in strength and number by the tight packing of side chains in the interface. All of this data supports a universal ice-binding mechanism that is emerging in which each AFP is shaped in such a way that a significant portion of its surface area can dock to ice. This interaction is specific to the ice surface that it binds and provides intimate surface-to-surface complementarity. Overall, the hydrogen bonds may serve to orient the protein on the ice, while hydrophobic residues protect these (*11*) and provide the van der Waals interactions that keep the protein bound and prevent the entrance of water molecules into the interface.

Clearly the dynamics of these side chains in the interface play an important role in the energetics of the interaction. We have addressed the side chain dynamics in two separate studies. Measured ^{13}C relaxation parameters for Cβ of the threonine residues indicated that these side chains experience a restricted range of motion on the picosecond to nanosecond time-scale of bond vector and librational motion. As well, we have shown that the threonine residues on the ice-binding face adopt a preferred χ_1 of -60° and that this preference increases as the temperature is lowered. The threonines not on the ice-binding face do not display this preferred conformation. This suggests that TmAFP adopts a rigid preformed ice binding structure in solution prior to recognition of the ice surface. Since the side chains do not have to reorient themselves prior to ice-binding, this reduces the entropic barrier for binding to a flat ice surface. We have also noted that the conformational preference of threonines on the TmAFP ice-binding face is lowered by mutation of neighbouring threonine residues to serine or valine. We suggest this is a result of an overall destabilization of the threonine array, which diminishes the necessary surface complementarity. The regularity of the threonine array is crucial for promoting ice binding. The extensive hydrogen bonding between successive β-strands helps stabilize the flat protein surface, prevents large-scale motions, and provides one of the dominant restrictions to motion of the ice-binding face.

Some implications of this work include experimental support for the hypothesis of surface complementarity as a key to the mechanism of ice binding by AFPs. We have shown that an indirect probe of the AFP-surface interaction, such as solution NMR spectroscopy, can provide detailed and valuable information to the study of AFP mechanism. Furthermore, the computational docking and molecular dynamics simulations which have been attempted on the AFP-ice interaction (*14*) will be improved by the inclusion of rigid side chains in the interface.

In terms of future work, I think the structure of an AFP bound to ice is an achievable goal and would provide great insight. Recent and ongoing work on solid-state NMR in our laboratory has shown that it is possible to obtain ^{13}C-CP/MAS spectra of a Type I AFP frozen in ice (*15*). This study indicates that the secondary structure remains intact upon freezing as the ^{13}Cα chemical shift of A17 remains at a value consistent with α-helical structure. However in this case, the AFP was flash frozen and completely overgrown in ice, and repeated cycles of freezing and thawing resulted in amyloid fibril formation. A look at an AFP frozen to ice in a physiological ice growth inhibition context using solid-state NMR would be invaluable.

Recent experiments describing the measurement of protein dynamics in supercooled water also interest me greatly (*16-18*). In these studies, the use of capillaries allows the sample to be supercooled to approximately –16°C (*17*). Even with the long overall rotational correlation time, heteronuclear NMR experiments could be successfully performed (*16*). This has allowed an examination of ^{15}N relaxation of supercooled ubiquitin at –13°C and revealed slow motional modes on the microsecond time-scale (*18*). It would be quite feasible to study AFPs under these conditions since they supercool very well. I have measured a one-dimensional spectrum of TmAFP at –12°C in a simple 5 mm O.D. medium-walled tube (Chapter 3) and sbwAFP has been measured to –8°C (*5*). I am certain that supercooling to at least –20°C would be achievable for these proteins and may yield further insight into their behaviour at these physiological temperatures.

Finally, I think comparison of the dynamics of more AFP types will be very informative. Studies with spruce budworm AFP have already confirmed my hypothesis that rigidity is important to ice-binding activity. Therefore, to show that fish AFPs indeed function by the same mechanism, demonstration of their rigidity would strengthen the conclusions already reached by the mutagenesis experiments.

References:

1. Cheng, C.C. (1998) *Curr. Opin. Gen. Dev.* **8**, 715-720.

2. Davies, P.L., Baardsnes, J., Kuiper, M.J., and Walker, V.K. (2002) *Phil. Trans. R. Soc. Lond. B* **357**, 927-935.

3. Fletcher, G.L., Hew, C.L., and Davies, P.L. (2001) *Annu. Rev. Physiol.* **63**, 359-390.

4. Baardsnes, J. and Davies, P.L. (2001) *Trends Biochem. Sci.* **26**, 468-469.

5. Graether, S.P., Gagné, S.M., Spyracopoulos, L., Jia, Z., Davies, P.L., and Sykes, B.D. (2003) *J. Mol. Biol.* **327**, 1155-1168.

6. Chao, H., Houston, Jr., M.E., Hodges, R.S., Kay, C.M., Sykes, B.D., Loewen, M.C., Davies, P.L., and Sönnichsen, F.D. (1997) *Biochemistry* **36**, 14652-14660.

7. Haymet, A.D.J., Ward, L.G., Harding, M.M., and Knight, C.A. (1998) *FEBS Lett.* **430**, 301-306.

8. Haymet, A.D.J., Ward, L.G., and Harding, M.M. (1999) *J. Am. Chem. Soc.* **121**, 941-948.

9. Zhang, W. and Laursen, R.A. (1998) *J. Biol. Chem.* **273**, 34806-34812.

10. Baardsnes, J., Kondejewski, L.H., Hodges, R.S., Chao, H., Kay, C., and Davies, P.L. (1999) *FEBS Lett.* **463**, 87-91.

11. Baardsnes, J. and Davies, P.L. (2002) *Biochim. Biophys. Acta* **1601**, 49-54.

12. Graether, S.P., Kuiper, M.J., Gagné, S.M., Walker, V.K., Jia, Z., Sykes, B.D., and Davies, P.L. (2000) *Nature* **406**, 325-328.

13. Marshall, C.B., Daley, M.E., Graham, L.A., Sykes, B.D., and Davies, P.L. (2002) *FEBS Lett.* **529**, 261-267.

14. Dalal, P., Knickelbein, J., Haymet, A.D.J., Sönnichsen, F.D., and Madura, J.D. (2001) *Phys. Chem. Commun.* **7**, 1-5.

15. Graether, S.P., Slupsky, C.M., and Sykes, B.D. (2003) *Biophys. J.* **84**, 552-557.

16. Skalicky, J.J., Sukumaran, D.K., Mills, J.L., and Szyperski, T. (2000) *J. Am. Chem. Soc.* **122**, 3230-3231.

17. Skalicky, J.J., Mills, J.L., Sharma, S., and Szyperski, T. (2001) *J. Am. Chem. Soc.* **123**, 388-397.

18. Mills, J.L. and Szyperski, T. (2002) *J. Biomol. NMR* **23**, 63-67.

Appendix A: Chemical shift assignments of TmAFP[*]

Table A.1: Resonance assignments for the *Tenebrio molitor* antifreeze protein

Residue	N (ppm)	HN (ppm)	Cα (ppm)	Hα (ppm)	Hβ (ppm)	Side chain (ppm)
1. Gln						
2. Cys			54.8	4.92		
3. Thr	114.1	7.80	60.5	5.29	4.14	Cβ - 70.4 Hγ* - 1.15
4. Gly	111.9	8.14		4.00 4.16		
5. Gly	103.2	7.71		4.01 4.21		
6. Ala	120.9	8.31	53.2	4.47	1.44	
7. Asp	115.1	8.12	55.0	4.95	2.60 2.92	
8. Cys	125.8	8.55	57.6	5.19	3.32 4.26	
9. Thr	122.1	8.85	55.3	4.67	3.64	Cβ - 69.1 Hγ* - 1.18
10. Ser	111.0	8.20	59.0	4.22	3.97 4.05	
11. Cys	120.7	7.91	57.8	4.44	3.63 3.71	
12. Thr	121.1	8.07	61.2	4.77	4.67	Cβ - 68.4 Hγ* - 1.16
13. Gly	112.0	7.37		3.52 4.63		
14. Ala	130.5	8.43	50.6	4.82	1.13	
15. Cys	126.6	8.70	55.3	5.34	3.64 3.73	
16. Thr	122.7	9.23	60.9	5.27	3.96	Cβ - 71.4 Hγ* - 1.13
17. Gly	114.8	8.25		3.89 4.29		
18. Cys	118.5	8.18	53.1	5.27	2.38 3.39	
19. Gly	109.2	8.40		3.88 4.28		
20. Asn	113.1	6.66	54.8	4.87	3.06 3.14	Nγ -111.4 $H\gamma_1$ - 6.63 $H\gamma_2$ - 7.44
21. Cys	123.1	8.40	56.2	5.21	3.38 3.70	

22. Pro			61.3	5.08	2.67	Hγ$_1$ - 2.64
					2.73	Hγ$_2$ - 2.68
						Hδ$_1$ - 3.64
						Hδ$_2$ - 3.73
23. Asn	115.2	8.96	52.9	5.03	2.43	Nγ - 116.8
					2.71	Hγ$_1$ - 7.27
						Hγ$_2$ - 8.13
24. Ala	124.6	6.76	53.5	4.02	1.66	
25. Val	115.6	8.08	63.6	4.34	2.23	Cβ - 72.5
						Hγ* - 1.07
26. Thr	113.4	7.74	61.2	4.97	3.97	Cβ - 70.8
						Hγ* - 1.19
27. Cys	120.9	8.19	55.3	5.20	3.44	
					3.64	
28. Thr	120.3	8.49	61.8	5.19	3.97	Cβ - 70.8
						Hγ* - 1.12
29. Asn	128.9	9.18	54.0	5.01	2.82	Nγ - 111.8
					3.10	Hγ$_1$ - 6.78
						Hγ$_2$ - 7.64
30. Ser	112.2	8.24	58.3	5.49	3.22	
					4.49	
31. Gln	116.8	7.53	54.0	5.15	1.33	Hγ* - 2.18
					1.88	
32. His	114.8	7.52		4.72	3.58	
					3.71	
33. Cys	119.7	7.80	58.9	4.84	3.00	
					3.84	
34. Val	111.8	7.09	65.0	4.32	2.37	Cβ - 71.7
						Hγ* - 1.04
35. Lys	121.9	8.68	55.9	4.39	1.60	Hγ* - 1.39
						Hδ* - 1.84
						Hε* - 3.06
36. Ala	125.1	6.92	53.1	4.16	1.31	
37. Asn	117.0	9.44	55.2	4.86	2.95	Nγ - 113.3
					3.02	Hγ$_1$ - 7.02
						Hγ$_2$ - 7.84
38. Thr	112.6	7.26	61.5	5.03	3.95	Cβ - 71.8
						Hγ* - 1.18
39. Cys	122.8	8.19	55.1	5.28	3.31	
					3.67	
40. Thr	119.6	8.64	61.5	5.43	4.02	Cβ - 71.7
						Hγ* - 1.24
41. Gly	122.1	8.68		4.23		
				4.67		

42. Ser	112.9	8.06	59.1	5.97	3.16 4.23	
43. Thr	109.3	8.04		4.71	4.35	Cβ - 70.2 Hγ* - 1.13
44. Asp	116.6	8.22	54.4	4.03	2.80	
45. Cys	118.3	7.52		4.72	2.76 3.55	
46. Asn	115.4	7.94	55.6	4.22	2.55 3.24	
47. Thr	106.7	9.13	65.9	4.92	4.35	Cβ - 69.2 Hγ* - 1.22
48. Ala	127.8	6.80	53.4	4.54	1.25	
49. Gln	118.3	8.94	56.8	4.52	2.26	Hγ* - 2.60
50. Thr	112.4	7.26	61.5	5.03	3.95	Cβ – 71.8 Hγ* - 1.18
51. Cys	123.5	8.51	54.9	5.13	3.18 3.76	
52. Thr	119.0	8.56	61.7	5.02	4.04	Cβ - 72.5 Hγ* - 1.24
53. Asn	130.4	9.30	54.0	5.29	2.76 3.16	Nγ - 110.8 Hγ_1 - 6.82 Hγ_2 - 7.80
54. Ser	111.3	8.39	58.1	5.72	3.13 4.01	
55. Lys	118.1	7.96	53.3	5.16	1.63	Hγ_1 - 1.32 Hγ_2 - 1.47 Hδ* - 1.75 Hε* - 2.90
56. Asp	118.9	8.22	55.1	3.68	1.82 2.35	
57. Cys	119.8	7.37	56.9	4.79	2.78 3.93	
58. Phe	116.2	6.54	54.9	4.68	3.28 3.68	Hδ* - 7.29 Hε* - 7.39 Hζ - 7.31
59. Glu	113.3	8.68	58.2	4.54	1.50 2.31	Hγ* - 2.49
60. Ala	125.3	6.96	54.0	4.41	1.28	
61. Asn	117.0	8.76	53.6	4.97	2.99	Nγ - 115.0 Hγ_1 - 7.10 Hγ_2 - 7.89
62. Thr	115.2	7.51	61.3	5.39	4.03	Cβ - 70.4 Hγ* - 1.24
63. Cys	124.6	9.13	55.0	5.38	3.21 3.32	

64. Thr	118.2	8.45	61.7	5.17	3.97	Cβ - 70.4 Hγ* - 1.22
65. Asp	132.4	8.89	56.1	5.06	2.68 3.15	
66. Ser	112.3	8.54	59.0	5.89	3.23 4.07	
67. Thr	108.9	7.96	58.2	5.25	4.08	Cβ - 68.4 Hγ* - 1.14
68. Asn	117.8	7.92	53.7	3.63	2.53 2.64	
69. Cys	116.6	8.16	56.9	4.78	3.01 3.66	
70. Tyr	114.6	6.60	55.9	4.82	3.00 3.29	Hδ* - 7.18 Hε* - 6.50
71. Lys	117.7	9.62	54.7	4.58	1.43 1.97	Hγ* - 1.63 Hδ* - 1.72 Hε* - 3.05
72. Ala	125.9	7.25	54.9	4.34	1.40	
73. Thr	114.5	8.14	60.7	4.52	4.29	Cβ - 71.0 Hγ* - 1.34
74. Ala	126.9	7.70		4.72	1.46	
75. Cys	126.1	8.81	55.7	5.09	2.65 3.57	
76. Thr	122.6	8.79	62.2	4.60	4.14	Cβ - 70.1 Hγ* - 1.32
77. Asn	123.1	7.85	54.2	4.52	3.12 3.32	
78. Ser	110.0	7.78	57.7	5.68	3.55 3.89	
79. Ser	117.1	8.19	57.2	4.83	3.97 4.19	
80. Gly	108.5	8.79		3.77 3.87		
81. Cys	117.6	7.94	55.0	4.97	2.96 3.58	
82. Pro			63.6	4.48	1.96 2.28	Hγ* - 1.88 $H\delta_1$ - 3.86 $H\delta_2$ - 3.95
83. Gly	110.3	8.59		3.93		
84. His						

---

[*] A version of this table has been published as supplementary material to Daley *et al.* (2002) *Biochemistry* **41**: 5515-5525 and has been deposited in the BMRB under accession number 5323.

Appendix B: Assessment of TmAFP folding by NMR

Citation:

Liou, Y.-C., Daley, M.E., Graham, L.A., Kay, C.M., Walker, V.K., Sykes, B.D., and Davies, P.L. (2000) Folding and structural characterization of highly disulfide-bonded beetle antifreeze protein produced in bacteria. *Protein Expression and Purification* **19**, 148-157.

Abstract:

The hyperactive antifreeze protein from the beetle, *Tenebrio molitor,* is an 8.5-kDa, threonine-rich protein containing 16 Cys residues, all of which are involved in disulfide bonds. When produced by *Escherichia coli,* the protein accumulated in the supernatant in an inactive, unfolded state. Its correct folding required days or weeks of oxidation at 22 or 4°C, respectively, and its purification included the removal of imperfectly folded forms by reversed-phase HPLC. NMR spectroscopy was used to assess the degree of folding of each preparation. One-dimensional ^{1}H and two-dimensional ^{1}H total correlation spectroscopy spectra were particularly helpful in establishing the characteristics of the fully folded antifreeze protein in comparison to less well-folded forms. The recombinant antifreeze protein had no free –SH groups and was rapidly and completely inactivated by 10 mM DTT. It had a thermal hysteresis activity of 2.5°C at a concentration of 1 mg/ml, whereas fish antifreeze proteins typically show a thermal hysteresis of ~1.0°C at 10–20 mg/ml. The circular dichroism spectra of the beetle antifreeze protein had a superficial resemblance to those of α-helical proteins, but deconvolution of the spectra indicated the absence of α-helix and the presence of β-structure and coil. NMR analysis and secondary structure predictions agree with the CD data and are consistent with a β-helix model proposed for the antifreeze protein on the basis of its 12-amino-acid repeating structure and presumptive disulfide bond arrangement.

NMR Contribution:

Materials and Methods – NMR spectroscopy

Samples for NMR spectroscopy were prepared by dissolving 4.5 mg protein (TmTHP isoform 2-14) in 500 μl 90% H_2O/10% D_2O (by vol), adjusting the pH to 5.5, and adding 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) as an NMR chemical shift reference. All experiments were carried out at 30°C on a Varian Unity 600 MHz spectrometer equipped with a pulse-field gradient triple-resonance probe. These experiments included two-dimensional ^{1}H homonuclear total correlation spectroscopy (TOCSY) spectra (*1, 2*) with a mixing time of 54 ms. The acquired data consisted of 4096 complex data points in the acquisition domain and 256 complex data points in the indirectly detected domain. The spectral width for both dimensions was 7500.5 Hz. The NMR spectra were processed using VNMR 5.3 software on a Sun workstation.

Results – NMR Analysis

The NMR spectroscopic data were of excellent quality and demonstrated that recombinant TmTHP is a stable, folded protein (Figure B.1). In the one-dimensional ^{1}H spectrum, the amide and aromatic protons at low-field are very well defined (Figure B.1A). A number of αCH protons are observed above and below the large residual water peak at approximately 4.9 ppm. All the methyl proton resonances are overlapped and occur upfield between 1.0 and 1.8 ppm, except for the N-terminal Met, which gives rise to the sharp peak at 2.2 ppm. This spectrum, along with the two-dimensional ^{1}H TOCSY spectrum in Figure B.1B, is typical of a high-resolution spectrum of a small, stable, and well-folded protein. As well, the wide chemical shift dispersion in the TOCSY spectrum for the proton resonances in individual amino acid spin systems indicates the folded nature of the protein. In general, the observation of downfield αCH protons resonating at 5.2–6.2 ppm and amide (NH) protons resonating at 8.0–9.6 ppm is characteristic of β-sheet secondary structure. Using the two-dimensional ^{1}H TOCSY spectrum and the protocol of Wishart *et al.* (*3*), the number of amide NH–αCH cross-peaks in the region defined by 7.1–10.0 and 4.85–6.2 ppm is half the number of residues involved in β-sheet and extended secondary structure. Thus for TmTHP, the 36 peaks in this region

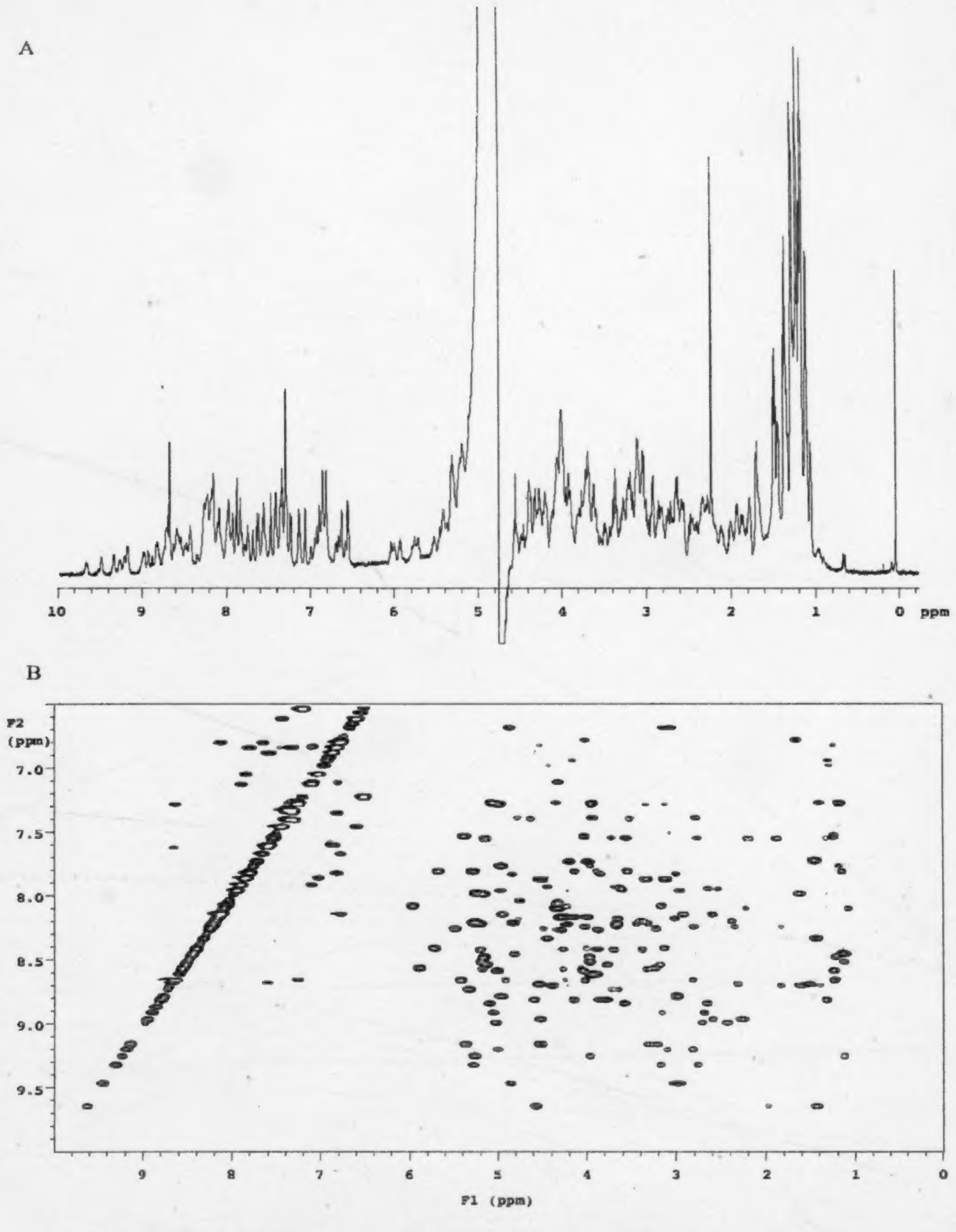


Figure B.1: ^1H NMR spectra of TmTHP. (A) One-dimensional ^1H NMR spectrum. The standard signal at 0 ppm is from the methyl proton resonances of DSS. The large residual peak at 4.9 ppm is due to H_2O. (B) Region of two-dimensional TOCSY ^1H NMR spectrum showing correlations of NH and aromatic protons to protons in the rest of the spectrum.

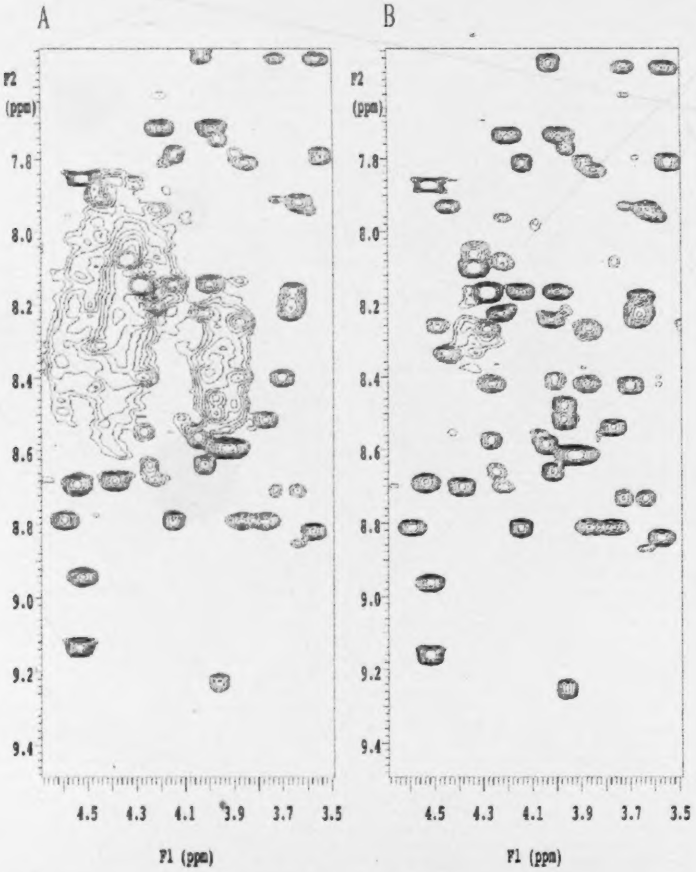


Figure B.2: Two-dimensional TOCSY NMR spectra. Comparative two-dimensional TOCSY NMR spectra focusing on the fingerprint region of the protein spectrum which shows correlations of NH protons to α-CH protons. This region is diagnostic of the misfolded TmTHP isoforms. The broad undefined region evident in the spectrum on the left (A) is clearly not present in the spectrum of the well-folded material (B).

correspond to 72 residues or approximately 85% β- and extended structure. The observed resonance linewidths are typical for an 8.5 kDa protein and therefore indicate that there is no aggregation of the protein in solution. Furthermore, there is no indication of any splitting or doubling of resonances that would be caused by incorrectly folded protein or by the protein existing in multiple conformations. Figure B.2 displays an expansion of the “fingerprint” region of the two-dimensional ^{1}H TOCSY spectrum, which shows some of the amide NH–αCH–βCH connections via the amino acid spin systems. The spectrum of the well-folded TmTHP displays strong, sharp peaks and clear individual correlations between protons (Figure B.2B). The spectrum in Figure B.2A is similar, displaying identical individual correlations, within an underlying broad, undefined region of NH–αCH connectivities from 7.8 to 8.6 and 4.3 to 4.5 ppm. This undefined region occurs only in this narrow chemical shift range, which is indicative of an unfolded random coil region in the protein. Unfolding in a portion of the protein leads to the loss of individual chemical environments that give rise to the usual chemical shift dispersion and results in the spreading out of the chemical shift and the observed broad, undefined region. It is not clear whether the unfolded region corresponds to complete unfolding of a certain percentage of all protein molecules or to unfolding in a particular region of each individual protein molecule. However, this misfolded form could largely be removed using an additional semipreparative reversed-phase HPLC step. In this way, ^{1}H NMR experiments were useful in quality control to ensure that the protein was properly folded.

References:

1. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson.* **53,** 521–528.

2. Davies, D. G., and Bax, A. (1985) *J. Am. Chem. Soc.* **107,** 2820–2821.

3. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991) *FEBS Lett.* **293,** 72–80.

Appendix C: Identification of the TmAFP ice-binding face

Citation:

Marshall, C.B., Daley, M.E., Graham, L.A., Sykes, B.D., and Davies, P.L. (2002) Identification of the ice-binding face of antifreeze protein from *Tenebrio molitor*. *FEBS Letters* **529**, 261-267.

Abstract:

The beetle *Tenebrio molitor* produces several isoforms of a highly disulfide-bonded β-helical antifreeze protein with one surface comprised of an array of Thr residues that putatively interacts with ice. In order to use mutagenesis to identify the ice-binding face, we have selected an isoform that folds well and is tolerant of amino acid substitution, and have developed a heating test to monitor refolding. Three different types of steric mutations made to the putative ice-binding face reduced thermal hysteresis activity substantially while a steric mutation on an orthogonal surface had little effect. NMR spectra indicated that all mutations affected protein folding to a similar degree and demonstrated that most of the protein folded well. The large reductions in activity associated with steric mutations in the Thr array strongly suggest that this face of the protein is responsible for ice binding.

NMR Contribution:

Materials and Methods – NMR spectroscopy

In order to determine the degree of folding of the wild-type TmAFP protein and the selected mutants, NMR spectra were obtained. Samples were prepared by dissolving approximately 2 mg of each protein in 500 μl of 90% H_2O/10% D_2O (by volume). The pH was adjusted to 5.3 with μl aliquots of 100 mM NaOD or DCl as required, and 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS) was added as an NMR chemical shift reference. All spectra were acquired at 30°C on a Varian Unity 600 MHz spectrometer equipped with a 5-mm triple resonance probe and *z*-axis pulsed field gradients. The experiments included two-dimensional ^{1}H homonuclear total correlation spectroscopy (TOCSY) spectra (*1*, *2*) with a mixing time of 54 ms. The acquired data consisted of 1024 complex data points in the acquisition domain and 256 complex data points in the indirectly detected domain. The spectral width for both dimensions was 7000.4 Hz. The NMR spectra were processed using VNMR 6.1B software on a Sun workstation.

Results and Discussion - Isoform Tm 4-9 folds better than 2-14

The folding of recombinant TmAFP isoform 2-14 takes several weeks and has an appreciable failure rate, with approximately one in four preparations yielding low TH activity. Because Tm 2-14 was difficult to refold, we compared the expression of another TmAFP isoform, Tm 4-9. Tm 4-9 protein is much more abundant than 2-14 in the hemolymph of *T. molitor* (*3*) and, although it only differs from 2-14 at five residues, recombinant Tm 4-9 was found to fold more rapidly and reliably than 2-14. The two-dimensional ^{1}H TOCSY spectra of both the Tm 2-14 and Tm 4-9 isoforms displayed in Figure C.1 are of excellent quality and are typical of high resolution spectra of small, stable and well-folded proteins. In particular, the wide chemical shift dispersion for the proton resonances in the individual amino acid spin systems indicates the folded nature of the protein. The observed resonance linewidths are typical for a protein of 8.5 kDa, suggesting there is no aggregation of the protein in solution. We have previously used two-dimensional ^{1}H TOCSY NMR spectra to identify improperly folded protein (*4*). In these spectra there is little indication of resonances that would be caused by incorrectly folded protein or by the protein existing in multiple conformations. Finally, the spectrum of the Tm 4-9 isoform is almost identical in chemical shift dispersion and resonance linewidth to that of Tm 2-14, which suggests that although the folding pathways may differ in rate and efficiency, the well-folded form of each isoform is ultimately very similar.

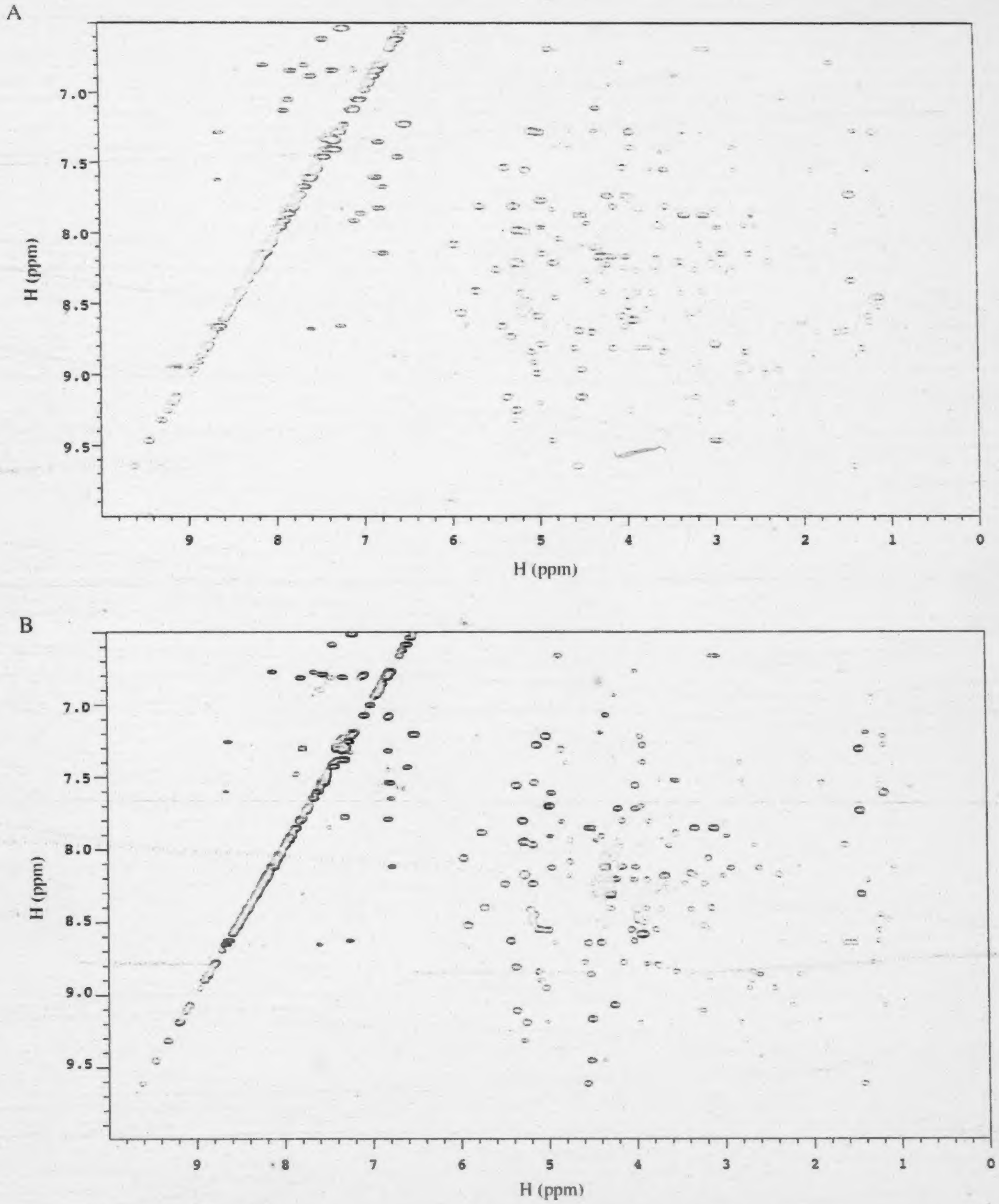


Figure C.1: Regions of the two-dimensional TOCSY ^{1}H NMR spectra showing correlations of NH and aromatic protons to protons in the rest of the spectrum for (A) TmAFP isoform 2-14 and (B) TmAFP isoform 4-9.

Results and Discussion – Tm 4-9 mutations[*]

Six steric mutations (T41Y, T41L, T41K, T63Y, T27Y, T39Y) were made to the putative ice-binding face to sample four different positions and three types of replacements at one position. The NMR spectra show that the wild-type protein (Figure C.2A), as well as the samples of two representative ice-binding face mutants (T27Y and T39Y, Figure C.2B and C), consist primarily of properly folded protein with a small fraction of unfolded protein present. This is evidenced by the existence of a poorly defined broad underlying component in the 'fingerprint' region of the two-dimensional ^{1}H TOCSY spectrum, which contains some of the amide NH-αCH-βCH connections via the amino acid spin systems. The broad, undefined region from 8.0 to 8.5 and 4.0 to 4.5 ppm is indicative of a small amount of unfolded, random coil protein (*4*). While reversed-phase HPLC is somewhat effective at separating well-folded TmAFP from improperly folded forms, it is difficult to remove all traces of unfolded protein.

To determine whether the loss of activity associated with steric mutations of Thr residues is specific to the putative ice-binding face, a Thr residue on a face of the protein that is orthogonal to the ice-binding face was also mutated to Tyr. The position T44 was chosen because this residue is part of the same loop as T41 and T39, which were also mutated to Tyr, and because this residue is surrounded by fewer bulky groups and is more exposed than the other Thr residues found outside the Thr array. The 'fingerprint' region of the two-dimensional ^{1}H TOCSY NMR spectrum (Figure C.2D) of this mutant is similar to the others shown, indicating that the sample also consists primarily of well-folded protein with a small percentage of unfolded protein. The mutation T44Y resulted in only ~20% loss of TH activity, suggesting that this Thr residue, and by inference the surrounding region of the β-helix, are not crucial for ice binding.

---

[*] Note the numbering of the TmAFP residues in this appendix are all one higher than I have used throughout the thesis. Member of the Davies' laboratory numbered the starting Met as 1 instead of 0 as I did.

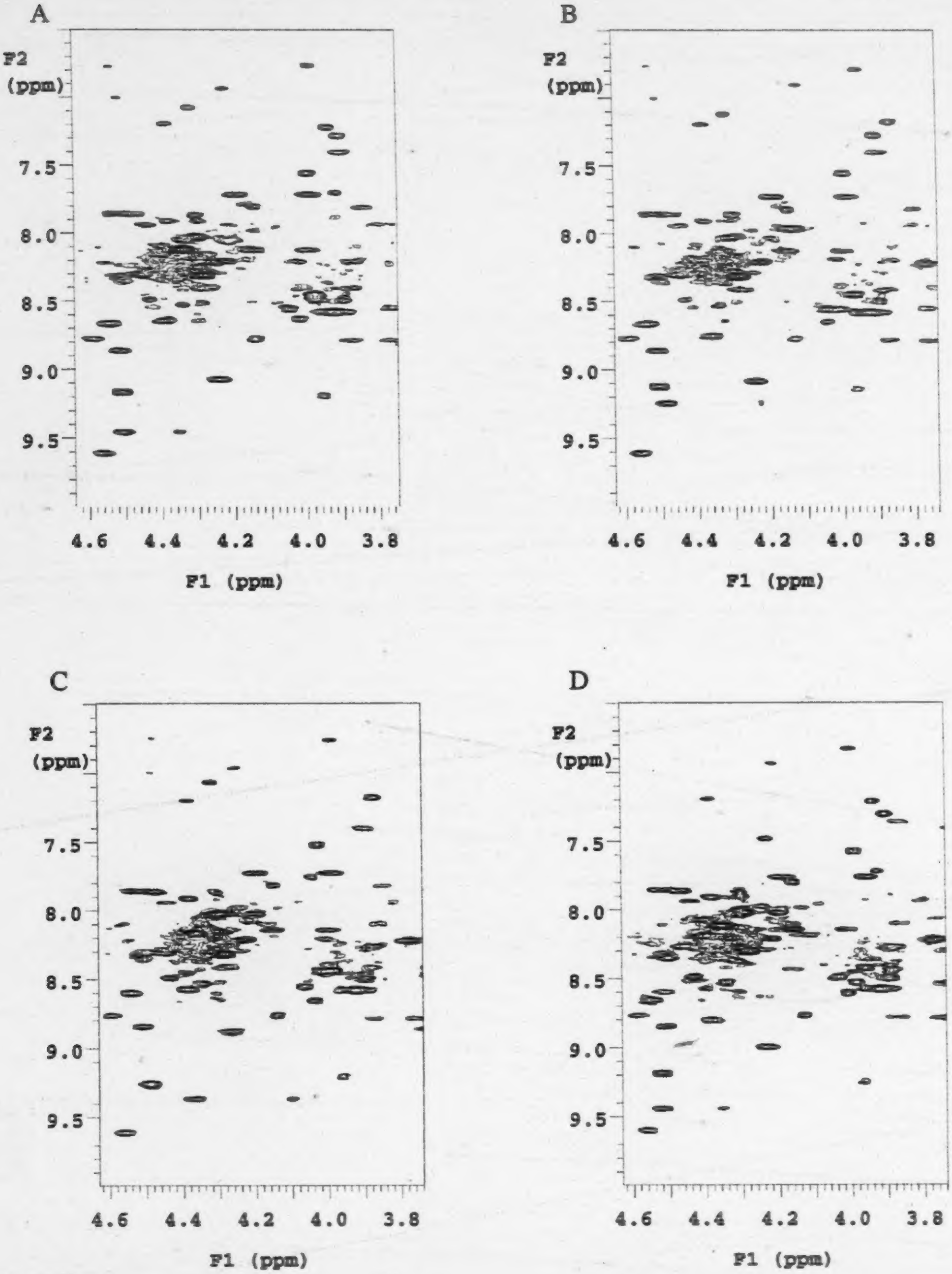


Figure C.2: Comparative two-dimensional TOCSY NMR spectra focusing on the fingerprint region of the protein spectrum, which shows correlations of NH protons to αCH (and some βCH) protons. This region is diagnostic of the misfolded TmAFP isoforms. (A) Wild-type TmAFP 4-9, (B) T27Y mutant, (C) T39Y mutant and (D) T44Y mutant.

NMR analysis confirmed that the incompletely folded form was a minor component of the purified AFP preparations and that the fraction of protein that was not well folded was fairly constant from one mutant to another. Therefore, it is reasonable to conclude that the ~90% loss of activity seen for the steric mutants T41Y and T39Y, versus the ~20% loss seen for the control mutation T44Y, reflects steric interference with the binding of the Thr array to ice.

References:

1. Braunschweiler, L. and Ernst, R.R. (1983) *J. Magn. Reson.* **53**, 521-528.

2. Davies, D.G. and Bax, A. (1985) *J. Am. Chem. Soc.* **107**, 2820-2821.

3. Liou, Y.-C., Thibault, P., Walker, V.K., Davies, P.L. and Graham, L.A. (1999) *Biochemistry* **38**, 11415-11424.

4. Liou, Y.-C., Daley, M.E., Graham, L.A., Kay, C.M., Walker, V.K., Sykes, B.D. and Davies, P.L. (2000) *Protein Expr. Purif.* **19**, 148-157.

Appendix D: Assessment of engineered TmAFP with deleted and added loops

Citation:

Marshall, C.B., Daley, M.E., Davies, P.L., and Sykes, B.D. Antifreeze activity varies with the number of coils in a β-helical insect antifreeze protein. In poster presentation at the PENCE Annual General Meeting, Edmonton, Alberta, May 10, 2002.

Abstract:

The thermal hysteresis (TH) activity of antifreeze proteins (AFPs) appears to correlate with the size of the protein. For example, an isoform of type I AFP from winter flounder with four 11-amino acid repeats is more active (on a molar or mg/mL basis) than the more common isoforms with only three repeats. To more thoroughly investigate the relationship between the size of an antifreeze protein and its activity, we have produced a series of recombinant proteins of various lengths based on the repetitive sequence of an AFP from the beetle *Tenebrio molitor* (Tm). Tm AFP isoforms are comprised of 7 to 10 tandem repeats of a 12-a.a. sequence (TCTxSxxCxxAx) that form internally disulfide-bonded coils that stack to form a right-handed β-helix. This fold presents a regularly-spaced array of Thr residues that is proposed to adsorb to ice. Through the addition or deletion of repeats, we have produced a series of proteins with 5 to 13 loops. Loss of one or more loops causes a drastic loss of TH activity whereas the addition of loops increases TH activity until a critical length is reached. We propose explanations why there is an upper limit to the enhancement of TH activity as a function of length.

NMR Contribution:

NMR spectra of Tm 4-9 with loop mutations were obtained to determine the degree of folding of each preparation. The samples were prepared by dissolving the protein in 500 μL of 90% H_2O / 10% D_2O (by volume). The pH was adjusted to approximately 6 with μL aliquots of 100 mM NaOD or DCl as required and 0.1 mM DSS was added as an NMR chemical shift reference. All spectra were acquired at 25°C on a Varian Unity 600 MHz spectrometer. 1D ^{1}H and 2D ^{1}H TOCSY spectra were collected as described previously (Appendices B and C). The 2D ^{1}H TOCSY spectra presented in the following pages represent in order:

Tm 4-9 minus 1 loop (72 residues)
Tm 4-9 wild type (84 residues)
Tm 4-9 plus 1 loop (96 residues)
Tm 4-9 plus 2 loops (108 residues)

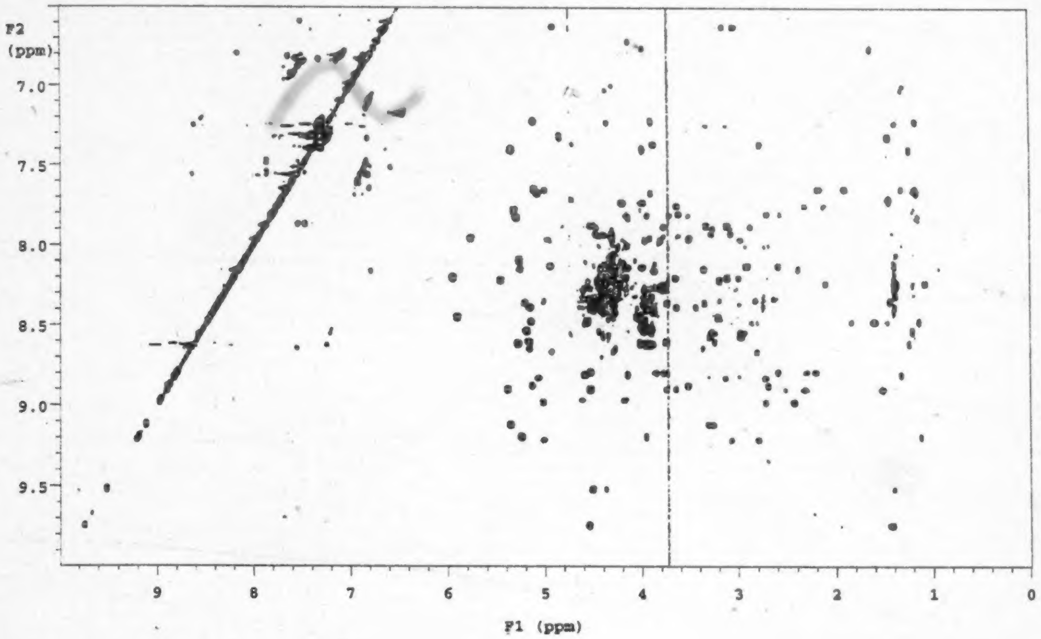
It is clear that the loop mutant preparations have some misfolded protein present, as evidenced by the broad component in the 'fingerprint' region of the spectra as we have observed previously. The proportion of misfolded protein appears to increase with the increasing size of the protein, as the plus 2 loop sample is quite poor compared to the others. However, it is clear that correctly folded protein is also present in each preparation despite the misfolded portion since there are also very sharp correlations with a wide chemical shift dispersion. In particular, the downfield αCH peaks from 5.4 – 6.2 ppm have been assigned to the serine residues that line the inside of the β-helix. In the wild-type protein there are five of these serines and five peaks are observed in this region of the spectrum. In the minus 1 loop spectrum, there are only four peaks in this region, indicating that one serine (and therefore one loop) is missing. In the plus 1 and plus 2 loop spectra, there are additional peaks in this region indicating that loops have been added and are properly folded.

There are two 1D ^{1}H spectra following the 2D ^{1}H TOCSY spectra. These are:

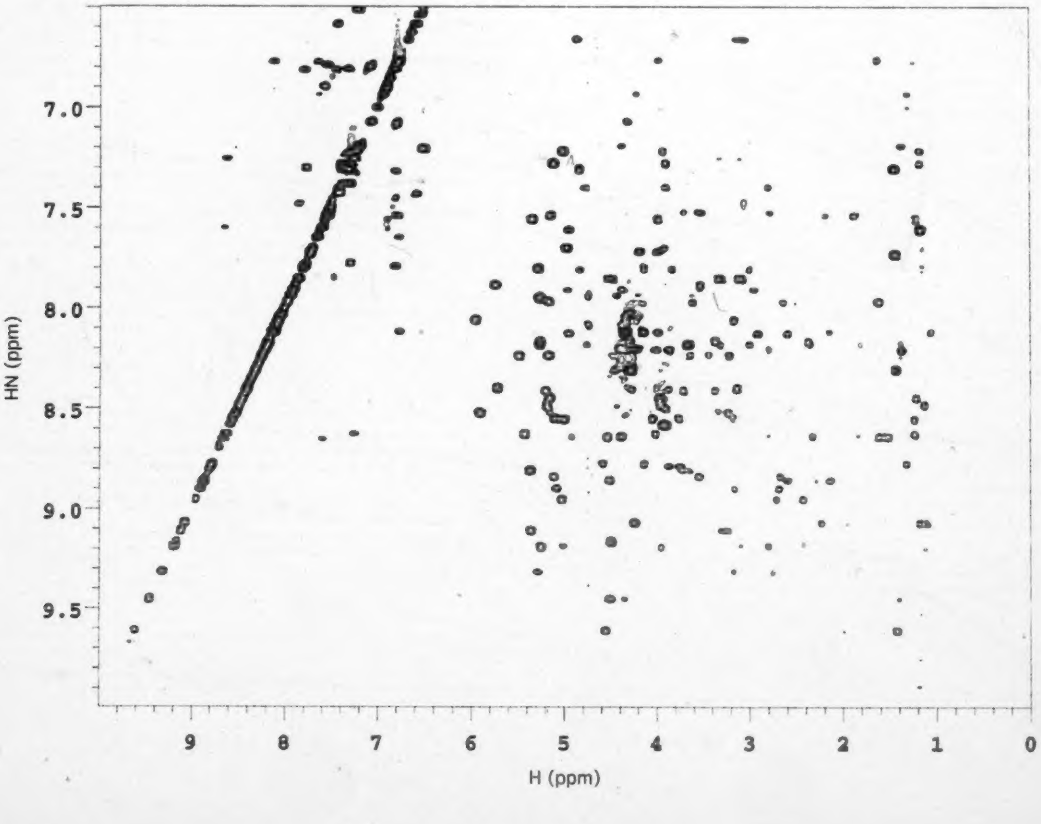
Tm 4-9 plus 3 loops (120 residues)
Tm 4-9 plus 4 loops (132 residues)

These spectra were of much poorer quality and only the 1D spectra were acquired. The plus 3 loop sample looks mostly unfolded, since there are no peaks in the amide region above 9 ppm. The plus 4 loop mutant sample appears to contain some well-folded protein. The chemical shift dispersion in this spectrum is better since there are amide peaks above 9 ppm and alpha peaks between 5.5 and 6 ppm. Overall, NMR spectroscopy proved to be very useful in assessing the folded state of the engineered TmAFP loop mutations.

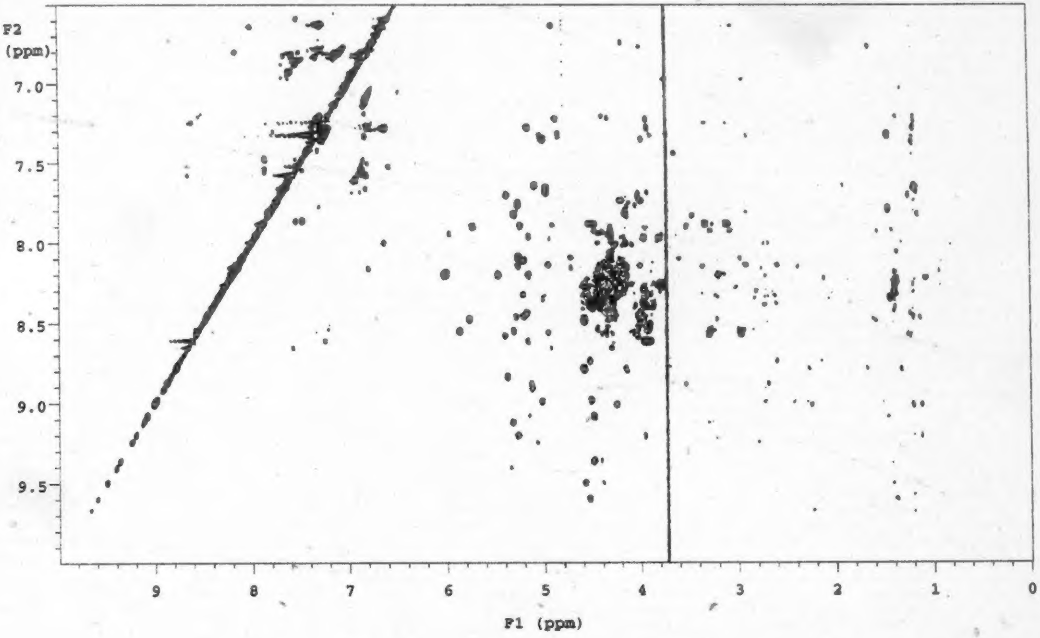
MINUS 1 LOOP



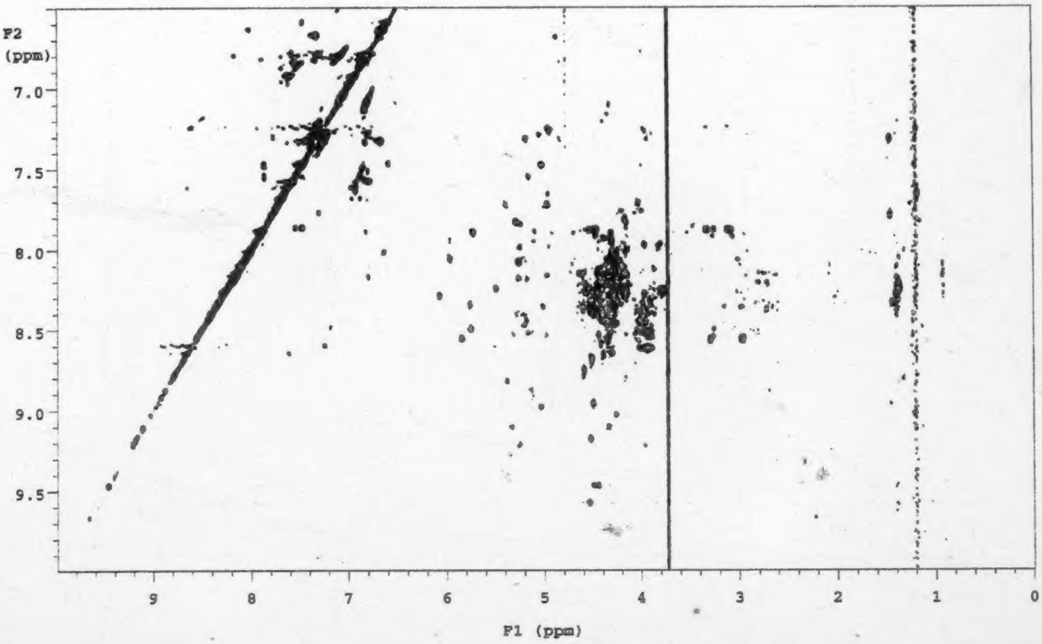
WILD TYPE



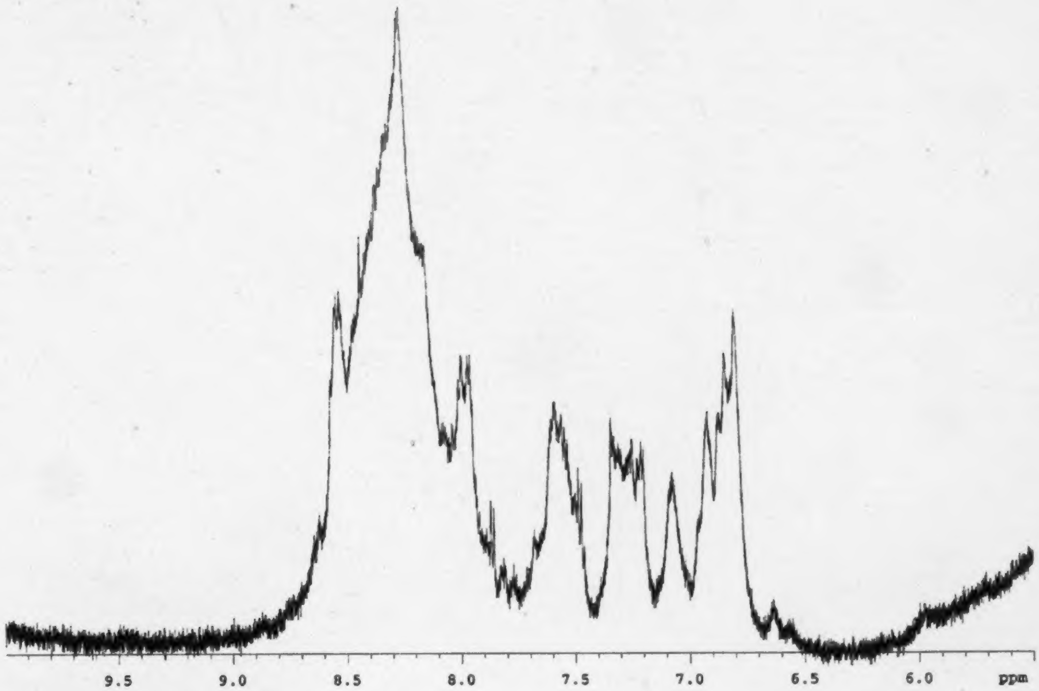
PLUS 1 LOOP



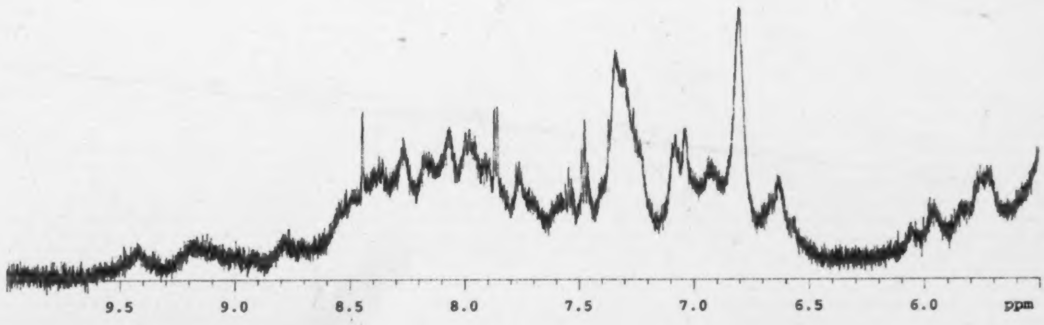
PLUS 2 LOOPS



PLUS 3 LOOPS



PLUS 4 LOOPS



Appendix E: Purification of TmAFP by inclusion into ice

Citations:

(1) Marshall, C.B., Kuiper, M.J., Lankin, C., Gauthier, S.Y., Walker,V.K., Tomczak, M.M., Daley, M.E., Sykes, B.D., and Davies, P.L. Partitioning of antifreeze proteins into ice. In poster presentation at the Symposium on Stress Proteins: From Antifreeze to Heat Shock, Bodega Bay, California, March 7-9, 2003.

(2) Marshall, C.B., Daley, M.E., Sykes, B.D., and Davies, P.L. Engineering of a repetitive β-helical antifreeze protein. In poster presentation at the PENCE Annual General Meeting, Vancouver, British Columbia, May 22-23, 2003.

Abstract:

From (2) - The beetle *Tenebrio molitor* (Tm) produces a number of antifreeze protein (AFP) isoforms, which are comprised of a variable number of 12-a.a. repeats (TCTxSxxCxxAx) that form internally disulfide-bonded coils and stack to form a right-handed β-helix. One side of the protein contains a regularly spaced array of Thr residues, which binds to ice, conferring antifreeze activity. A number of naturally occurring AFP isoforms as well as engineered AFP fusion proteins suggest that larger AFPs possess higher antifreeze activity. We have exploited the repetitive sequence and structure of Tm AFP to investigate the relationship between the size and activity of an antifreeze protein. Using the 7-loop isoform Tm 4-9 as a template, we have produced a series of recombinant proteins from 5 to 13 loops, by adding and deleting internal repeats. We were able to significantly increase antifreeze activity by adding loops, up to a critical maximum length, whereas deletion of loops caused a marked loss of activity. NMR was used to assess the folding of each variant. The preparations of the larger proteins contained more misfolded material, presumably because with each added loop, one extra disulfide bond must be correctly formed. **Because misfolded forms of TmAFP are not easily removed by chromatography, we have used a novel purification method, based on the affinity of active AFPs for ice, to separate well-folded protein from misfolded forms. This method has proved to be the critical step for accurately determining specific activities of well-folded proteins.**

NMR Contribution:

A novel method for purifying antifreeze proteins (AFPs) by virtue of their affinity for, and inclusion into, ice (*1*) has recently been described. As has been described, AFPs bind to the surface of ice. At high concentrations, they inhibit ice growth, while at low concentrations, they bind but are overgrown and get included into the ice. This inclusion is exploited in the ice-etching technique used to determine ice-binding planes, which was described in the Introduction to this thesis. Non-AFPs and solutes are almost entirely excluded from slowly growing ice.

Ice affinity purification (IAP) can be used to remove the misfolded form of TmAFP. As discussed in the previous appendices, NMR is very useful as a quality control method to determine the folded state of AFP preparations. In Appendix D, the folding of various loop mutations was assessed. One of the long isoforms of TmAFP, with two additional 12 amino acid repeats consists of a mixture of well-folded and misfolded protein by NMR (as in Appendix D). It is very difficult to remove the misfolded material, even with reversed phase HPLC. However this was accomplished using IAP as NMR spectroscopy was able to confirm. The NMR spectra clearly show that the misfolded protein was excluded from the ice while the well-folded material was included in the ice fraction.

This allowed us to resolve the question we had posed in our previous work on TmAFP folding (Appendix B). We had not been able to determine whether the unfolded region corresponded to complete unfolding of a certain percentage of all protein molecules or to unfolding in a particular region of each individual protein molecule. This method clearly indicates that it is indeed a certain percentage of the protein population that is completely unfolded, and that this misfolded protein can be removed.

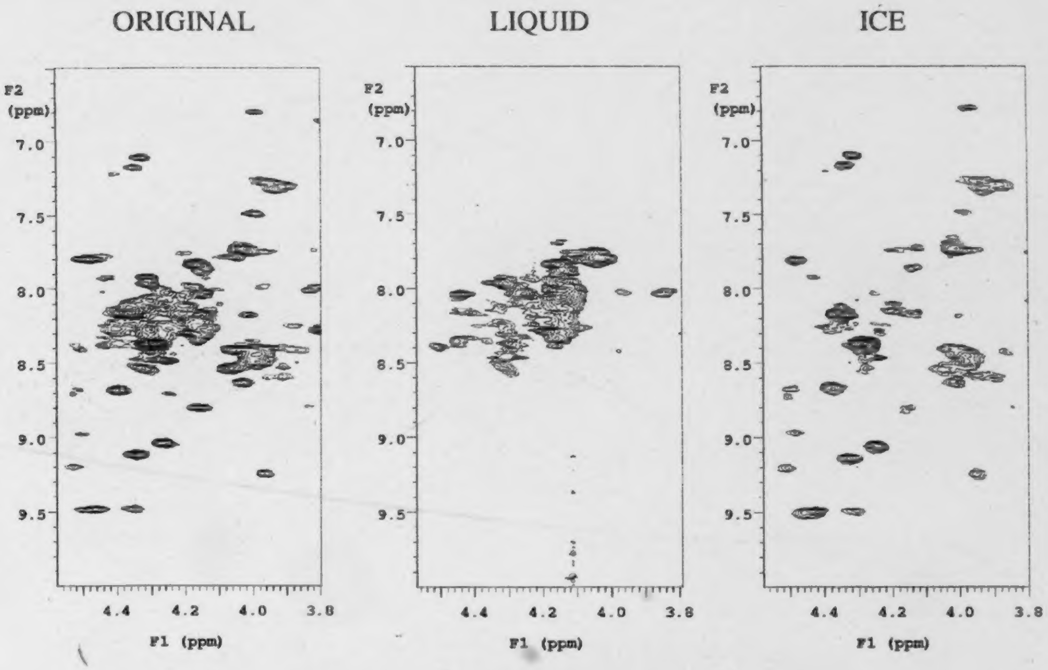


Figure E.1: Two-dimensional TOCSY ^{1}H NMR spectra of the 'fingerprint region' showing correlations of NH protons to α-CH protons. The original sample shows clear individual correlations with an underlying broad, undefined region. The ice-affinity purified material shows primarily the clear sharp peaks indicating well-folded material, whereas the liquid fraction contains the broad peaks indicative of unfolded random coil.

References:

1. Kuiper, M.J., Lankin, C., Gauthier, S.Y., Walker, V.K. and Davies, P.L. (2003) *Biochem. Biophys. Res. Commun.* **300**, 645-648.